

FILE COPY

AR0 23010.4-MA



TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843-3143

2

Department of STATISTICS
Statistical Interdisciplinary
Research Laboratory
E41SEP@TAMU.MI.BITNET

Emanuel Parzen
Distinguished Professor

**A UNIFIED APPROACH
TO ESTIMATING TAIL BEHAVIOR**

Technical Report #55

May 1989

AD-A210 961

Scott D Grimshaw

DTIC
ELECTE
AUG 02 1989
S B D

Texas A&M Research Foundation

Project No. 5641

Sponsored by the U. S. Army Research Office

Professor Emanuel Parzen, Principal Investigator

Approved for public release; distribution unlimited

89

185

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARO 23010.4-MA	2. GOVT ACCESSION NO. N/A	3. RECIPIENT'S CATALOG NUMBER N/A
4. TITLE (and Subtitle) A Unified Approach to Estimating Tail Behavior		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Scott D Grimshaw		8. CONTRACT OR GRANT NUMBER(s) DAAL03-87-K-0003
9. PERFORMING ORGANIZATION NAME AND ADDRESS Texas A&M University Department of Statistics College Station, TX 77843-3143		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE May 1989
		13. NUMBER OF PAGES 116
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) NA		
18. SUPPLEMENTARY NOTES The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) quantile data analysis, extreme value distributions, pareto distributions, tail estimation, exceedances over threshold		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Tail estimators are proposed which make minimal assumptions and let the data dictate the form of the probability model. These estimators use only the observations in the tail and are based on a unifying density-quantile model. The fundamental result in this work is a representation of the quantile function of the exceedances over a threshold. This representation (1) motivates a unified parameterization for tail estimators of the underlying probability model; (2) motivates methods for obtaining parameter estimates; and (3) simplifies the derivation of the asymptotic properties of the proposed parameter estimates.		

ABSTRACT

A Unified Approach to Estimating Tail Behavior. (May 1989)

Scott D Grimshaw, B.S., Southern Utah State College;

M.S., Texas A&M University

Chair of Advisory Committee: Dr. Emanuel Parzen

Tail estimators are proposed which make minimal assumptions and let the data dictate the form of the probability model. These estimators use only the observations in the tail and are based on a unifying density-quantile model. The fundamental result in this work is a representation of the quantile function of the exceedences over a threshold. This representation (1) motivates a unified parameterization for tail estimators of the underlying probability model; (2) motivates methods for obtaining parameter estimates; and (3) simplifies the derivation of the asymptotic properties of the proposed parameter estimates.

Parameter estimates may be obtained using a Generalized Pareto Distribution or a Generalized Extreme Value Distribution model of the exceedences. Assuming the underlying distribution can be correctly classified as either short tailed or long tailed, other estimates are formed. The asymptotic properties of these estimates are derived under rate of convergence conditions to show the effect of threshold selection on parameter properties. $(k \rightarrow \infty) \leftarrow$

The parameters are shown to be nonidentifiable and their estimators contain a bias which may approach zero very slowly. Therefore, if the parameters are the focus of the analysis, extremely large sample sizes are required to reduce the bias to a negligible amount. If the tail estimates are of interest, the bias is less likely to be serious and the nonidentifiability problem provides a closer approximation to the tail for small samples.



Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

ACKNOWLEDGEMENTS

My formal education as a statistical scientist has been most gratifying. Many have played a significant part in my development and it is a pleasure to acknowledge their role.

I consider myself most fortunate for the opportunity to research under Professor Parzen's direction. I gratefully acknowledge his role in my development as a researcher and thank him for his support and guidance. I thank him for teaching me to courageously investigate intuition and to appreciate and strive for the aesthetic dimension of statistical science where beauty and utility are united.

I am grateful to all those who have contributed to my understanding of probability and statistics. To Jim Cotts I owe a great debt for introducing me to statistics and planting the seeds for my future. Joe Newton, Mike Longnecker, Ron Hocking, and Tom Wehrly stand out during these past six years of graduate school at Texas A&M. My attitudes toward teaching follow from these pedagogical role models.

I thank Don Marquardt and Bill Smith for creating the DuPont-Texas A&M Statistics Intern Program and choosing me as the first intern in the program. Learning to solve real problems with statistics from the ASG statisticians added an essential dimension to my education and renewed my enthusiasm for statistics.

To my fellow graduate students I am indebted for their discussions and friendship. I wish to thank my early office mates Alan Zimmermann and Major Ron Berdine for setting high standards for my remaining graduate school years. I am also grateful to a statistical brother, Will Alexander, for a delightful professional relationship over the past few years. Further, I thank Will and his wife Cathy for a friendship which made the long frustrating days seem shorter and the moments of success more satisfying.

Throughout my research I have relied heavily on the computing knowledge of Will Alexander and Joe Newton. I am further indebted to Will for MATRIX and Joe for TIMESLAB, both of which were used to create the graphs and perform the computations in this dissertation.

TABLE OF CONTENTS

	Page
ABSTRACT	
DEDICATION	
ACKNOWLEDGEMENTS	
TABLE OF CONTENTS	
LIST OF TABLES	
LIST OF FIGURES	
1. INTRODUCTION	1
2. EXCEEDENCE OVER THRESHOLD APPROACH TO TAIL ESTIMATES	7
2.1. A Unifying Model For Tail Behavior	7
2.2. Conditional Distribution Of The Exceedence Over A Threshold	10
2.3. Tail Estimates Based On Exceedences Over A Threshold	12
2.4. Expected Values For Functions Of The Exceedences	16
3. PARAMETER ESTIMATES FROM THE GENERALIZED PARETO DISTRIBUTION	18
3.1 Method	18
3.2 Parameters Estimated By Maximum Likelihood	19
3.3 Parameters Estimated From The Sample Quantile Process	29
4. PARAMETER ESTIMATES FROM THE GENERALIZED EXTREME VALUE DISTRIBUTION	32
4.1 Method	32
4.2 Parameters Estimated By Maximum Likelihood	33
4.3 Parameters Estimated From The Sample Quantile Process	39
5. PARAMETER ESTIMATES FROM MODELS BASED ON REGULARLY VARYING EXPRESSIONS FOR THE TAILS	41
5.1 Parametric Modeling For The Tail	41
5.2 Estimation Assuming The Class Of Tail Behavior Is Short Tailed	47
5.3 Estimation Assuming The Class Of Tail Behavior Is Long Tailed	55

TABLE OF CONTENTS (continued)

	Page
6. DISCUSSION OF PARAMETER ESTIMATE PROPERTIES . . .	61
6.1 Comparison Of Parameter Estimates	61
6.2 Interpretation Of Parameter Estimates	62
6.3 Effect Of Bias In Parameter Estimates On Tail Estimates .	65
7. THRESHOLD SELECTION	67
7.1 Optimal Thresholds Based On The Parameters	67
7.2 Optimal Thresholds Based On The Tail	68
8. MOTIVATING EXAMPLE	69
8.1 Tail Estimates For The Feather River	76
8.2 Tail Estimates For The Blackstone River	76
9. CONCLUDING REMARKS	79
REFERENCES	81
APPENDIX A	
COMMON PARAMETRIC PROBABILITY MODELS	85
APPENDIX B	
PROOFS OF SECTION 2.3 THEOREMS	97
APPENDIX C	
PROOF OF THEOREM 2.4.1	99
APPENDIX D	
GRADIENT AND HESSIAN OF THE GPD LOG-LIKELIHOOD .	102
APPENDIX E	
GRADIENT AND HESSIAN OF THE GEV LOG-LIKELIHOOD .	103
APPENDIX F	
DERIVATIVES OF HALL'S ESTIMATING EQUATIONS	106
VITA	107

LIST OF TABLES

Table	Page
1. Common parametric probability models expressed as tail behavior models where $fQ(1-u) = u^{\rho+1}L(u)$	9
2. Failure rate of the Newton-Raphson optimization in two dimensions of the log-likelihood of the GPD, reproduced from Hosking and Wallis (1987)	27
3. Quantile functions of common parametric probability models expressed in the representation given in Theorem 5.1.1(a) based on the tail behavior model $fQ(1-u) = u^{\rho+1}L(u)$	43
4. Distribution functions of common parametric probability models expressed in the representation given in Theorem 5.1.1(b) based on the tail behavior model $fQ(1-u) = u^{\rho+1}L(u)$	44
5. Probability density functions of common parametric probability models expressed in the representation given in Theorem 5.1.1(c) based on the tail behavior model $fQ(1-u) = u^{\rho+1}L(u)$	45
6. Annual floods (1902-1960) for the Feather River at Oroville, California taken from Benjamin and Cornell (1970)	70
7. Annual floods (1929-1965) for the Blackstone River at Woonsocket, Rhode Island taken from Wood, Rodríguez-Iturbe, and Schaake (1974)	71
8. Table containing the optimal threshold percentile and parameter estimates for the proposed tail estimates based on the exceedences of a threshold for the Feather River annual floods	77
9. Table containing the optimal threshold percentile and parameter estimates for the proposed tail estimates based on the exceedences of a threshold for the Blackstone River annual floods	78

LIST OF FIGURES

Figure	Page
1. Estimated $F(x)$, $Q(u)$, and $f(x)$ when a sample of $n = 20$ from a symmetric unimodal probability model is treated as a sample from a normal distribution (solid line with blocks) and a Cauchy distribution (dotted line)	3
2. Graph of the GPD log-likelihood function for a generated random sample	24
3. Graph of the function $h(\eta)$ used to simplify the search for the zeroes of the gradient vector in the two-dimensional parameter space for a GPD random sample	26
4. Identification Quantile Plots for the Uniform ($\rho < 0$), Normal, and Cauchy ($\rho > 0$) Distributions clearly differentiating the types of tail behavior	46
5. Graph of the quantile function for the Normal and Lognormal distributions	64
6. Identification quantile box plots for the data used as examples	72
7. Graph of the estimated Gumbel model (solid line), Lognormal model (dotted line), and Pearson Type III model (solid line with blocks) overlaid on the sample quantile function (step function) for the data used as examples	74
8. Graph of the estimated Gumbel model (solid line), Lognormal model (dotted line), and Pearson Type III model (solid line with blocks) overlaid on the sample quantile function (step function) on the upper quartile for the data used as examples	75
9. Graph of the proposed tail estimates based on the exceedences of a threshold for the Feather River annual floods on the upper quartile	77
10. Graph of the proposed tail estimates based on the exceedences of a threshold for the Blackstone River annual floods on the upper quartile	78
11. Uniform Distribution	86

LIST OF FIGURES (continued)

Figure	Page
12. Negative Exponential Distribution	87
13. Negative Weibull(p) Distribution	88
14. Exponential Distribution	89
15. Logistic Distribution	90
16. Normal Distribution	91
17. Weibull(p) Distribution	92
18. Lognormal Distribution	93
19. Cauchy Distribution	94
20. Pareto(p) Distribution	95
21. Fréchet(p) Distribution	96

1. INTRODUCTION

Suppose that the possible observed values from a population can be characterized by a random variable X whose probability model is estimated using a sample from the population. The properties of this estimated probability model which correspond to the population characteristics of interest are the foundation of statistical analysis.

Three important functions of a probability model for a continuous random variable are the absolutely continuous distribution function $F(x)$, the quantile function $Q(u)$, and the density function $f(x)$. The significance of these three functions in statistical analysis follows from their interpretation as key properties of the population.

For example, the distribution function $F(x)$ is the probability that an observed value from the population will be less than or equal to a given value of x , i.e. $F(x) = P[X \leq x]$. In applications where the observed values are times until failure or death, the distribution function for a given value of x is the probability that the lifetime will be less than or equal to x . A more optimistic expression of this information is the survival function used in reliability. The probability that the lifetime will exceed a given value of x is $S(x) = P[X > x] = 1 - F(x)$.

The quantile function $Q(u)$ is the smallest value of x such that the probability of a value greater than or equal to x is equal to u , i.e. $Q(u) = F^{-1}(u) = \inf\{x : F(x) = u\}$. In applications, the quantile function is used to determine the value of x such that an observed value of this magnitude (or greater) occurs with probability u for a given value of u .

The density function $f(x)$ of an absolutely continuous distribution function represents the probability X is in the interval (a, b) for $a < b$ as the area under the density function between a and b , i.e. $P[a < X < b] = \int_a^b f(x)dx$. The density function is used to describe many properties of the population graphically. Characteristics such as modality and skewness are evident from plots of $f(x)$.

In some applications, the population characteristics of primary interest cor-

The format and style follows that of *The Annals of Statistics*.

respond to the tails of the distribution function, quantile function, and density function. For example, an experimenter investigating the lifetime of a product will want the probability of an early death or an exceptionally long life. That is, the value of the distribution function $F(x)$ for values of x with $F(x)$ near zero or values of x with $F(x)$ near one. A hydrologist analyzing an annual flood record will want the magnitude of rare high level floods. That is, the value of the quantile function $Q(u)$ for values of u near one. An experimenter may investigate the tails of the density function $f(x)$ to graphically display the concentration of possible values at the extremes.

This work focuses on the problem of estimating the tails of $F(x)$, $Q(u)$, and $f(x)$ from a random sample. The most basic estimators of $F(x)$ and $Q(u)$ from a sample of size n are the sample distribution function defined as

$$F_n(x) = \left\{ \text{fraction of the observed values less than or equal to } x \right\}, \quad x \in \mathbb{R},$$

and the sample quantile function defined as

$$Q_n(u) = \left\{ [nu + 1] \text{th largest observed value} \right\}, \quad 0 < u < 1,$$

where $[\cdot]$ denotes the greatest integer operation. Nonparametric density estimators follow this same vein as basic estimators of the density function.

The sample distribution function, sample quantile function, and nonparametric density estimate are typically used in early stages of statistical analysis since they make minimal assumptions on the underlying probability model. These estimators are important data analytic tools used as other known characteristics of the population are incorporated to formulate other estimates.

The classical approach to tail estimation is to assume the underlying probability model belongs to some known class \mathcal{P} whose elements are indexed by a parameter θ taking values in a set Θ , i.e. $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$. The distribution function, quantile function, and density function then have parametric representations $F(x; \theta)$, $Q(u; \theta)$, and $f(x; \theta)$. Tail estimates are given by $F(x; \hat{\theta})$, $Q(u; \hat{\theta})$, and $f(x; \hat{\theta})$, where $\hat{\theta}$ denotes an estimate of the parameter θ based on the sample from the population.

The beauty of this classical parametric approach is tarnished by what Fisher (1948) called the problem of specification. Often it is difficult to select a single parametric family for the population. Several candidates may appear reasonable judging from their fit to the observed values.

To demonstrate this complication, suppose that a random sample is taken from a population characterized by a symmetric unimodal probability model. Two possible parametric families are the normal and the Cauchy. Figure 1 contains graphs of estimated $F(x)$, $Q(u)$, and $f(x)$ when a sample of $n = 20$ from a symmetric unimodal probability model are treated as a sample from a normal distribution and a Cauchy distribution. Both estimates are overlaid on the sample distribution function, sample quantile function, and a kernel density estimator.

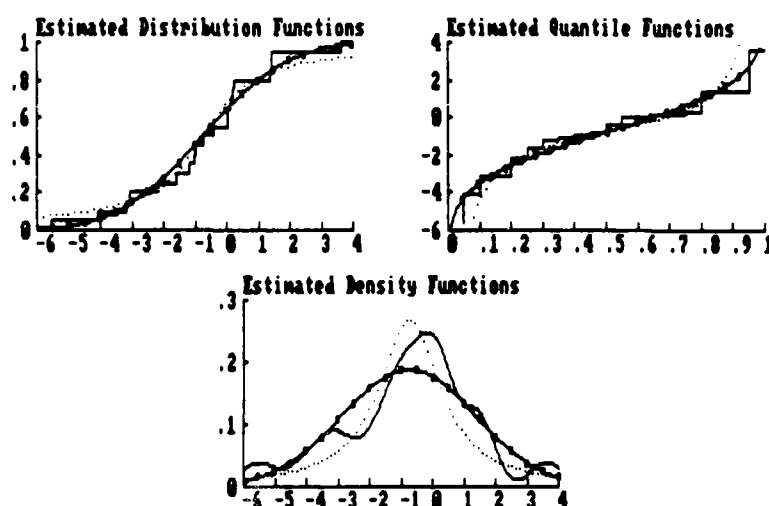


FIG. 1. Estimated $F(x)$, $Q(u)$, and $f(x)$ when a sample of $n = 20$ from a symmetric unimodal probability model is treated as a sample from a normal distribution (solid line with blocks) and a Cauchy distribution (dotted line). Estimates are overlaid on the sample distribution function, sample quantile function, and a kernel estimate of the density function (solid line). The normal and Cauchy modeling lead to very different tail inference despite yielding similar inference for central values.

Notice that the two parametric estimators yield similar inference for central values of the random variable. However, the focus of this work is on tail values, not central values, and inference at the tails is quite different under the two

parametric models. Extremely small and extremely large values are much more likely under the Cauchy modeling. The distribution function $F(x)$ approaches zero and one much more rapidly under the normality assumption. The quantile function $Q(u)$ for the Cauchy model decreases more rapidly in a neighborhood of zero and increases more rapidly in a neighborhood of one. The density function $f(x)$ for the Cauchy model has much more area in the tail.

It is very difficult to discriminate between the different possible parameterizations even when the possible parametric models specify very different tail properties. In this example, the sample size is too small for a goodness of fit test to have sufficient power to detect differences in the observed tail and the fitted tail under the normal and Cauchy modeling. The tails of the sample distribution function, sample quantile function, and nonparametric density estimates have insufficient observations in the tail to indicate important properties of tail behavior.

This work proposes estimators of $F(x)$, $Q(u)$, and $f(x)$ which are applicable under minimal assumptions. These estimators can be used in applications where little is known about the underlying population. The estimators can also be used in a data analytic sense to validate tail behavior properties in probability modeling applications. The work is outlined as follows.

Section 2 proposes the model for tail behavior, defines tail behavior parameters, and summarizes the characteristics of these parameters. The model for tail behavior is a basic result from which two approaches to tail estimation can be unified. Generally applicable tail estimates are proposed using only those observations which exceed a threshold value, i.e. the observations in the tail.

The fundamental result of this work is stated in this section. The quantile function for the exceedences can be represented as the sum of a function which can be parameterized and a deterministic error function demonstrating the dependence on the threshold value. This representation motivates a parametric tail estimation model, motivates methods for obtaining parameter estimates, and simplifies the derivation of asymptotic properties of the proposed parameter estimates.

Three approaches to the problem of parameter estimation are considered. The first two treat the exceedences as a random sample from a parametric family motivated from the representation for the quantile function of the exceedences. Section 3 investigates the Generalized Pareto distribution (GPD) modeling and Section 4 investigates the Generalized Extreme Value (GEV) distribution modeling.

An innovative approach to tail parameter estimates using the ideas of continuous parameter time series on the quantile process is introduced. These ideas are stimulated from the work of Parzen (1979) on location and scale parameter estimates.

The most popular choice for parameter estimates is maximum likelihood. A new algorithm is proposed for the numerical computation of the GPD maximum likelihood estimates. This algorithm corrects the inadequacies of common Newton-Raphson type algorithms.

The second approach to tail estimates follows from representations which are derived from the general tail behavior model. Section 5 proposes parameter estimates based on the largest order statistics assuming a parametric model is valid beyond the threshold. The properties of these estimators are treated in two cases since the parameterization for the tail depends on a prior assumption on the tail behavior.

A comparison of the different parameter estimates is made in Section 6. All the estimators are shown to be biased, and no global statements can be made regarding an 'optimal' estimator. A popular use of the parameter estimates is as diagnostics for existence of variance and higher order moments. However, great caution must be exercised in interpreting parameter estimates for reasons given in this section.

Section 7 discusses the important question of threshold selection. In order to reduce the bias, the threshold must be chosen as large as possible. However, this reduces the number of observations used in the estimators and inflates the variance of the estimates. A threshold selection procedure is proposed which minimizes the distance between the estimated distribution function and the sample

distribution function over the tail values.

A motivating example is provided in Section 8. The data for this example considers the problem of estimating the tail of the quantile function for two rivers from a history of observed annual floods. The high dependency on the choice of parametric family is demonstrated. The tail estimators proposed in this work are applied as alternative estimators which make minimal assumptions on the underlying probability model.

Concluding remarks are made in Section 9.

2. EXCEEDENCE OVER THRESHOLD APPROACH TO TAIL ESTIMATES

2.1. A Unifying Model For Tail Behavior

2.1.1. *Notation.* Let X_1, \dots, X_n be a random sample from a population with strictly increasing, absolutely continuous distribution function $F(x)$, density function $f(x) = F'(x)$, *quantile function*

$$Q(u) = F^{-1}(u) = \inf\{x : F(x) \geq u\}, \quad 0 \leq u \leq 1,$$

density-quantile function $fQ(u) = f \circ Q(u)$, and *quantile density function* $q(u) = Q'(u)$. Notice that $fQ(u) \cdot q(u) = 1$.

From the random sample, define the sample distribution function

$$F^n(x) = \frac{i}{n}, \quad X(i; n) \leq x < X(i+1; n), \quad i = 0, 1, \dots, n,$$

and the sample quantile function

$$Q^n(u) = X(i; n), \quad \frac{i-1}{n} < u \leq \frac{i}{n}, \quad i = 1, \dots, n,$$

where $X(i; n)$ denotes the i th order statistic in the random sample of size n , $X(0; n) = -\infty$, and $X(n+1; n) = \infty$.

2.1.2. *Tail Behavior Model.* Parzen (1979) has suggested that the behavior of the density-quantile function $fQ(u)$ in the neighborhood of $u = 0$ and $u = 1$ can be used to classify the tail behavior of a probability model. The classification of any continuous probability model follows from expressing

$$(2.1.1) \quad fQ(1-u) = u^{\rho+1}L(u),$$

where ρ is called the right tail exponent of the probability model and $L(u)$ is a slowly varying function as $u \rightarrow 0^+$, i.e. $L(u)$ is a positive measurable function defined on $[0, \infty)$ satisfying

$$\lim_{u \rightarrow 0^+} \frac{L(\lambda u)}{L(u)} = 1 \quad \text{for all } \lambda > 0.$$

Table 1 contains examples of common parametric probability models from Appendix A expressed according to (2.1.1).

An associated left tail exponent can be defined also. However, this work considers only the right tail without loss of generality since applications to the left tail can be made by negating the random variable.

The tail exponent ρ is finite if and only if X_1 has a finite moment of order δ for some $\delta > 0$. In this research, only continuous probability models where (2.1.1) holds with finite ρ are considered. However, this is not an all inclusive family. For example, a random variable with distribution function $F(x) = 1 - (\ln x)^{-1}$, $x > e$ has $\rho = \infty$ (and hence no finite moments).

Estimating the tail exponent has become popular because $\rho < 1/\delta$ if and only if $E|X_1|^\delta < \infty$. In particular, testing $H_0 : \rho < \frac{1}{2}$ is used as a diagnostic for finite variance. Other work on tail exponent estimation expresses the distribution function as $F(x) = 1 - x^{-\alpha}L^*(x)$, where $\alpha > 0$ and $L^*(x)$ is a slowly varying function as $x \rightarrow \infty$. It is shown in Section 5 that these two parameterizations for the tail exponent satisfy $\rho\alpha = 1$ for $\alpha, \rho > 0$.

The use of slowly varying functions in defining the tail exponent is just one application of the concept introduced in 1930 by J. Karamata as a suitable class of functions in connection with a Tauberian theorem for Laplace transforms. Bingham, Goldie, and Teugels (1987) review the generalization to regularly varying functions and provide examples of applications to probability theory in the areas of stability and domains of attraction, central limit theory, renewal theory, queues, occupation times, and extreme value theory.

Examples of slowly varying functions as $u \rightarrow 0^+$ include:

- (i) any positive measurable functions with positive limits at zero; for example, $L(u) = \Delta[1 + O(r(u))]$ where $\Delta > 0$ and $r(\cdot)$ is a positive measurable function with $\lim_{u \rightarrow 0^+} r(u) = 0$;
- (ii) $L(u) = -\ln u$;
- (iii) $L(u) = \ln \ln \cdots (-\ln u)$;
- (iv) $L(u) = \exp\left\{[-\ln u]^{-\alpha_1} [\ln(-\ln u)]^{\alpha_2} \cdots [\ln \ln \cdots \ln(-\ln u)]^{\alpha_k}\right\}$, where k is a positive integer and $0 < \alpha_i < 1$ for $i = 1, \dots, k$;

TABLE 1

Common parametric probability models expressed as tail behavior models where $fQ(1-u) = u^{\rho+1}L(u)$. The parameter ρ is the tail exponent and $L(u)$ is a slowly varying function as $u \rightarrow 0^+$. In some cases, an asymptotically equivalent expression for $L(u)$ is given which is in the form of the slowly varying function examples.

Distribution	Density-Quantile Function
Uniform	$fQ(1-u) = u^{-1+1} \cdot 1$
Neg. Exponential	$fQ(1-u) = u^{-1+1} \cdot (2 \ln 3)(1-u)$
Neg. Weibull(p)	$fQ(1-u) = u^{(-1/p)+1} \cdot \sigma p(1-u)[-u^{-1} \ln(1-u)]^{(-1/p)+1}$ $\sim u^{(-1/p)+1} \cdot \sigma p[1 - .5(3 - p^{-1})u]$ as $u \rightarrow 0^+$
Exponential	$fQ(1-u) = u^{0+1} \cdot (2 \ln 3)$
Logistic	$fQ(1-u) = u^{0+1} \cdot (4 \ln 3)(1-u)$
Normal	$fQ(1-u) = u^{0+1} \cdot \sigma \phi[\Phi^{-1}(1-u)]/u$ $\sim u^{0+1} \cdot \sigma(-2 \ln u)^{1/2}$ as $u \rightarrow 0^+$
Weibull(p)	$fQ(1-u) = u^{0+1} \cdot \sigma p(-\ln u)^{(-1/p)+1}$
Lognormal	$fQ(1-u) = u^{0+1} \cdot \sigma \phi[\Phi^{-1}(1-u)]/u \cdot e^{-\Phi^{-1}(1-u)}$ $\sim u^{0+1} \cdot \sigma(-2 \ln u)^{1/2} \cdot e^{-\Phi^{-1}(1-u)}$ as $u \rightarrow 0^+$
Cauchy	$fQ(1-u) = u^{1+1} \cdot (4/\pi)[\sin^2 \pi(1-u)]/u^2$ $\sim u^{1+1} \cdot 4\pi[1 - (\pi^2/3)u^2]$ as $u \rightarrow 0^+$
Pareto(p)	$fQ(1-u) = u^{(1/p)+1} \cdot \sigma p$
Fréchet(p)	$fQ(1-u) = u^{(1/p)+1} \cdot \sigma p(1-u)[-u^{-1} \ln(1-u)]^{(1/p)+1}$ $\sim u^{(1/p)+1} \cdot \sigma p[1 - .5(3 + p^{-1})u]$ as $u \rightarrow 0^+$

Note: See Appendix A for the definition of σ , a different scale constant for each different distribution.

$$(v) L(u) = \exp\left\{[-\ln u]/[\ln(-\ln u)]\right\}$$

Further examples of slowly varying functions can be generated from those given by noting two properties of slowly varying functions:

- (a) if $L(u)$ varies slowly as $u \rightarrow 0^+$, so does $[L(u)]^\alpha$ for all $\alpha \in \mathbb{R}$;
- (b) if $L_1(u)$ and $L_2(u)$ vary slowly as $u \rightarrow 0^+$, so do $L_1(u) \cdot L_2(u)$ and $L_1(u) + L_2(u)$. Further, if $L_2(u) \rightarrow 0$ as $u \rightarrow 0^+$, then $L_1 \circ L_2(u)$ is also slowly varying as $u \rightarrow 0^+$.

Some of the tail behavior representations given in Table 1 contain asymptotically equivalent expressions for $L(u)$ as examples of the formulations for slowly varying functions and the properties given above.

2.2. Conditional Distribution Of The Exceedence Over A Threshold

This research proposes estimates of the tails of the distribution function, density function, and quantile function for the family of random variables with finite ρ . These estimators use only the exceedences over a high threshold value. This approach allows the observed values in the tail to dictate the tail estimate.

The exceedence over a threshold is denoted by $X - T$ given $X > T$ for a given threshold T satisfying $Q(0) < T < Q(1)$. It is easily shown that the exceedences have distribution function

$$F_{X-T | X>T}(x; T) = \frac{F(T+x) - F(T)}{1 - F(T)}, \quad x > 0.$$

This distribution function expression is used by other authors to derive properties of tail estimates and tail exponent estimates.

However, this work suggests a representation for the quantile function of the exceedences which

- (1) motivates a unified parameterization for the tail of $F(x)$, $f(x)$, and $Q(u)$;
- (2) motivates parameter estimates; and
- (3) simplifies the derivation of the asymptotic properties of the parameter estimates and the tail estimates for $F(x)$, $f(x)$, and $Q(u)$.

Before stating this representation in the following theorem, define the *hazard*

quantile function as

$$hQ(u) = h \circ Q(u) = \frac{f \circ Q(u)}{1 - F \circ Q(u)} = \frac{fQ(u)}{1 - u},$$

and the power transformation (also called the Box-Cox (1964) transformation) for $z > 0$ and $\lambda \in \mathbb{R}$ as

$$g(z; \lambda) = \begin{cases} \frac{z^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln z, & \lambda = 0. \end{cases}$$

THEOREM 2.2.1. Suppose that $fQ(1 - u) = u^{\rho+1}L(u)$, where $\rho \in \mathbb{R}$ and $L(u)$ is slowly varying as $u \rightarrow 0^+$. Then

$$(2.2.1) \quad Q_{X-T | X>T}(u; T) = \frac{1}{hQ(1 - t^*)} \left[-g(1 - u, -\rho) + \epsilon(t^*, 1 - u, \rho) \right]$$

where $t^* = 1 - F(T)$ and

$$\epsilon(t, u, \rho) = \int_u^1 z^{-\rho-1} \left[\frac{L(t)}{L(tz)} - 1 \right] dz.$$

The proof of this theorem and those that follow in this subsection are given in Appendix B.

The representation given by (2.2.1) expresses the quantile function for the exceedences as the sum of two functions. The first does not depend on the threshold and motivates a parametric model for the quantile function of the exceedences based on the tail exponent ρ . The deterministic error function $\epsilon(t, u, \rho)$ expresses the systematic bias of this parameterization.

The convergence of $\epsilon(t, u, \rho)$ to zero as $T \rightarrow Q(1)^-$ is an important property. The following theorem states the uniform convergence and a rate of convergence result for the deterministic error function.

THEOREM 2.2.2. Suppose that $fQ(1 - u) = u^{\rho+1}L(u)$, where $\rho \in \mathbb{R}$ and $L(u)$ is slowly varying as $u \rightarrow 0^+$.

(a) Then, for every $0 < \delta < 1$, $\lim_{T \rightarrow Q(1)^-} \epsilon(t^*, u, \rho) = 0$ uniformly in $\delta \leq u < 1$, where $t^* = 1 - F(t)$.

(b) Further suppose that

$$\left| \frac{L(t)}{L(tu)} - 1 \right| \leq A(u)R(t)$$

for some positive measurable functions $A(u)$ and $R(t)$ where $\lim_{t \rightarrow 0^+} R(t) = 0$. Then there exists a positive measurable function $A^*(u)$ such that $|\epsilon(t, u, \rho)| \leq A^*(u)R(t)$.

Previous work with the exceedences assumes $T \rightarrow Q(1)^-$ which permits the simplification with $\epsilon(t^*, 1 - u, \rho) = 0$. The effect of the rate at which the threshold converges is revealed in the generalization in this work to rates of convergence for $\epsilon(t, u, \rho)$. For example, the most popular expression for $L(u)$ in the exceedence literature is $L(u) = \Delta[1 + O(u^\gamma)]$, for $\Delta > 0$, $\gamma > 0$. In this case, $T \rightarrow Q(1)^-$ such that $R(t^*) = (t^*)^\gamma \rightarrow 0$ where $t^* = 1 - F(T)$. However, the thresholds may be required to converge much more rapidly. If, for example, $L(u) = -\ln u$, then $T \rightarrow Q(1)^-$ satisfying $R(t^*) = -1/\ln t^* \rightarrow 0$.

2.3. Tail Estimates Based On Exceedences Over A Threshold

An important use of the representation (2.2.1) for the quantile function of the exceedences is the parametric model suggested for the tails of the quantile function, distribution function, and density function of the underlying population.

To begin motivating this parameterization, first notice that the hazard quantile function $hQ(\cdot)$ is used in (2.2.1) as a standardization, but any positive measurable function $a(\cdot)$ satisfying $\lim_{t \rightarrow 0^+} a(t) \cdot hQ(1 - t) = 1$ may replace $hQ(1 - t)$. This gives the more general expression

$$(2.3.1) \quad Q_{X-T | X > T}(u; T) = a(t^*)[-g(1 - u; -\rho) + \epsilon(t^*, 1 - u, \rho)]$$

where $t^* = 1 - F(T)$. The effect on the rate of convergence result is that $|\epsilon(t^*, 1 - u, \rho)| \leq A^*(u)R^*(t^*)$ if $a(t)hQ(1 - t) = 1 + O(R_1(t))$, where $R^*(t) = \max\{R(t), R_1(t)\}$.

The quantile function of the underlying population can be expressed by unconditioning the quantile function of the exceedences. The expression for the distribution function follows by inversion, and from that, the density function expression follows by differentiation. Hence, the tails can be written as

$$(2.3.2) \quad Q(u) = T + a(t^*) \left[-g \left(\frac{1-u}{t^*}; -\rho \right) + \epsilon_Q(t^*, u, \rho) \right]$$

for $1 - t^* < u < 1$

$$(2.3.3) \quad F(x) = 1 - t^* \cdot g^{-1} \left(-\frac{1}{a(t^*)} (x - T); -\rho \right) + \epsilon_F(t^*, x, \rho)$$

for $T < x < Q(1)$

$$(2.3.4) \quad f(x) = t^* \frac{1}{a(t^*)} \cdot (g^{-1})' \left(-\frac{1}{a(t^*)} (x - T); -\rho \right) + \epsilon_f(t^*, x, \rho)$$

for $T < x < Q(1)$

where

$$g^{-1}(z, \lambda) = \begin{cases} (1 + \lambda z)^{1/\lambda}, & \lambda < 0, z < 0 \\ e^z, & \lambda = 0, z < 0 \\ (1 + \lambda z)^{1/\lambda}, & \lambda > 0, -1/\lambda < z < 0 \end{cases}$$

and

$$(g^{-1})'(z, \lambda) = \begin{cases} (1 + \lambda z)^{(1/\lambda)-1}, & \lambda < 0, z < 0 \\ e^z, & \lambda = 0, z < 0 \\ (1 + \lambda z)^{(1/\lambda)-1}, & \lambda > 0, -1/\lambda < z < 0 \end{cases}$$

It is easy to show that if $|\epsilon(t, u, \rho)| \leq A^*(u)R^*(t)$ for some positive measurable functions $A^*(u)$ and $R^*(t)$ where $\lim_{t \rightarrow 0^+} R^*(t) = 0$, then $|\epsilon_Q(t, u, \rho)| \leq A_Q^*(u)R^*(t)$, $|\epsilon_F(t, x, \rho)| \leq A_F^*(x) \cdot t R^*(t)$, and $|\epsilon_f(t, x, \rho)| \leq A_f^*(x) \cdot t R^*(t)$ for some positive measurable functions $A_Q^*(u)$, $A_F^*(x)$, and $A_f^*(x)$.

A parameterization for the tail of $Q(u)$, $F(x)$, and $f(x)$ can be motivated by assuming the functions $\epsilon_Q(t, u, \rho) = 0$, $\epsilon_F(t, u, \rho) = 0$, $\epsilon_f(t, u, \rho) = 0$ and treating ρ and $a = a(t^*)$, a scalar given the threshold, as parameters. Sections 3-5 propose different parameter estimates for the tail exponent ρ and the scaling parameter a which have not previously been considered under a unified theory.

The paradigm for estimating the tails of $Q(u)$, $F(x)$, and $f(x)$ from a random sample is as follows:

1. From a random sample X_1, \dots, X_n , choose, as a function of n , a threshold percentile t_n close to zero.
2. Estimate the corresponding threshold $T_n = Q^-(1 - t_n)$.
3. Obtain parameter estimates $(\hat{\rho}, \hat{a})$ from the exceedences $X_i - Q^-(1 - t_n)$ for all $X_i > Q^-(1 - t_n)$.
4. Estimate the tails of the quantile function, distribution function, and density function by

$$(2.3.5) \quad Q^{\wedge}(u) = Q^-(1 - t_n) + \hat{a} \left[-g \left(\frac{1-u}{t_n}; -\hat{\rho} \right) \right] \quad \text{for } 1 - t_n < u < 1,$$

$$(2.3.6) \quad F^{\wedge}(x) = 1 - t_n \cdot \left[g^{-1} \left(-\frac{1}{\hat{a}} [x - Q^-(1 - t_n)]; -\hat{\rho} \right) \right] \\ \text{for } Q^-(1 - t_n) < x < Q(1),$$

$$(2.3.7) \quad f^{\wedge}(x) = t_n \frac{1}{\hat{a}} \cdot (g^{-1})' \left(-\frac{1}{\hat{a}} [x - Q^-(1 - t_n)]; -\hat{\rho} \right) \\ \text{for } Q^-(1 - t_n) < x < Q(1).$$

The estimates for ρ and a proposed in Sections 3-5 will be shown to have asymptotically normal distributions given T_n as $nt_n^* \rightarrow \infty$, where $t_n^* = 1 - F(T_n)$. Therefore, the asymptotic normality of the tail estimates follows since they are functions of asymptotically normal random variables.

THEOREM 2.3.1. *Suppose that conditional on T_n , with $t_n = 1 - F(T_n)$,*

$$\begin{bmatrix} \hat{\rho}_n \\ \hat{a}_n \end{bmatrix} \quad \text{is} \quad \text{AN} \left(\begin{bmatrix} \rho_0 \\ a_0 \end{bmatrix}, (nt_n^*)^{-1} \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} \right)$$

as $nt_n^ \rightarrow \infty$ for some $\rho_0 \neq 0$, $a_0 > 0$, and scalars v_{ij} such that the covariance matrix is positive definite.*

(a) *For $1 - t_n^* < u < 1$,*

$$Q^{\wedge}(u) \quad \text{is} \quad \text{AN} \left(T_n + a_0 \cdot [-g((1-u)/t_n^*; -\rho_0)], (nt_n^*)^{-1} \cdot \sigma_Q^2(u) \right),$$

as $nt_n^* \rightarrow \infty$, where

$$\begin{aligned}
 \sigma_Q^2(u) = & \frac{1}{\rho_0^2} \left\{ v_{11} a_0^2 \left\{ \frac{1}{\rho_0} - \left(\frac{1-u}{t_n^*} \right)^{-\rho_0} \left[\frac{1}{\rho_0} + \ln \left(\frac{1-u}{t_n^*} \right) \right] \right\}^2 \right. \\
 & + (v_{12} + v_{21}) a_0 \left[\left(\frac{1-u}{t_n^*} \right)^{-\rho_0} - 1 \right] \\
 & \cdot \left\{ \frac{1}{\rho_0} - \left(\frac{1-u}{t_n^*} \right)^{-\rho_0} \left[\frac{1}{\rho_0} + \ln \left(\frac{1-u}{t_n^*} \right) \right] \right\} \\
 & \left. + v_{22} \left[\left(\frac{1-u}{t_n^*} \right)^{-\rho_0} - 1 \right]^2 \right\}
 \end{aligned}
 \tag{2.3.8}$$

(b) For $T_n < x < Q(1)$,

$$F^*(x) \quad \text{is} \quad \text{AN} \left(1 - t_n^* \cdot g^{-1}(-(1/a_0)(x - T_n); -\rho_0), (t_n^*/n) \cdot \sigma_F^2(x) \right),$$

as $nt_n^* \rightarrow \infty$, where

(2.3.9)

$$\begin{aligned}
 \sigma_F^2(x) = & \frac{(x - T_n)^2}{\rho_0^4 a_0^4} \left[1 + \frac{\rho_0}{a_0} (x - T_n) \right]^{-2/\rho_0} \\
 & \cdot \left\{ v_{11} a_0^2 \left[\ln \left(1 + \frac{\rho_0}{a_0} (x - T_n) \right) \right]^2 \right. \\
 & + (v_{12} + v_{21}) \rho_0^2 a_0 \left[1 + \frac{\rho_0}{a_0} (x - T_n) \right]^{-1} \ln \left(1 + \frac{\rho_0}{a_0} (x - T_n) \right) \\
 & \left. + v_{22} \rho_0^4 \left[1 + \frac{\rho_0}{a_0} (x - T_n) \right]^{-2} \right\}
 \end{aligned}$$

(c) For $T_n < x < Q(1)$,

$$f^*(x) \quad \text{is} \quad \text{AN} \left(t_n^* (1/a_0) \cdot (g^{-1})'(-(1/a_0)(x - T_n); -\rho_0), (t_n^*/n) \cdot \sigma_f^2(x) \right),$$

as $nt_n^* \rightarrow \infty$, where

(2.3.10)

$$\begin{aligned}
\sigma_f^2(x) = & \frac{1}{\rho_0^4 a_0^4} \left[1 + \frac{\rho_0}{a_0} (x - T_n) \right]^{-2[(1/\rho_0)+1]} \\
& \cdot \left\{ v_{11} (x - T_n)^2 \left[\ln \left(1 + \frac{\rho_0}{a_0} (x - T_n) \right) \right]^2 \right. \\
& + (v_{12} + v_{21}) \rho_0^2 (x - T_n)^2 \ln \left(1 + \frac{\rho_0}{a_0} (x - T_n) \right) \\
& \cdot \left\{ (x - T_n) \left(\frac{\rho_0 + 1}{a_0} \right) \left[1 + \frac{\rho_0}{a_0} (x - T_n) \right]^{-1} - 1 \right\} \\
& \left. + v_{22} \rho_0^4 \left\{ (x - T_n) \left(\frac{\rho_0 + 1}{a_0} \right) \left[1 + \frac{\rho_0}{a_0} (x - T_n) \right]^{-1} - 1 \right\}^2 \right\}.
\end{aligned}$$

2.4. Expected Values For Functions Of The Exceedences

In determining the properties of estimators proposed in Sections 3-5, the expectation of certain functions of the exceedences are needed. In the quantile domain, moments are easily found noting the relation

$$E_{X-T | X>T} [\psi(X-T)] = \int_0^1 \psi[Q_{X-T | X>T}(u)] du$$

for any function $\psi(\cdot)$ where $E_{X-T | X>T} |\psi(X-T)| < \infty$.

The proof of the following theorem uses this quantile expression for expectation, the representation for the quantile function of the exceedences, and the rate of convergence result for the deterministic error function.

THEOREM 2.4.1. *Suppose that $fQ(1-u) = u^{\rho+1}L(u)$, where $\rho \in \mathbb{R}$, $\rho \neq 0$ and*

$$\left| \frac{L(t)}{L(tu)} - 1 \right| \leq A(u)R(t)$$

for some positive measurable functions $A(u)$ and $R(t)$ where $\lim_{t \rightarrow 0+} R(t) = 0$. For a given threshold value $Q(0) < T < Q(1)$, let $t^ = 1 - F(T)$. Further, let $a = a(t^*)$ be the scalar value of a function $a(\cdot)$ satisfying $a(t^*)hQ(1-t^*) =$*

$1 + O(R_1(t^*))$ for some positive measurable function $R_1(t)$ with $\lim_{t \rightarrow 0^+} R_1(t) = 0$. Also, suppose $-\rho[Q(1) - T]hQ(1 - t^*) = 1 + O(R_2(t^*))$ for some positive measurable function $R_2(t)$ with $\lim_{t \rightarrow 0^+} R_2(t) = 0$. Finally, suppose $\rho T hQ(1 - t^*) = 1 + O(R_3(t^*))$ for some positive measurable function $R_3(t)$ with $\lim_{t \rightarrow 0^+} R_3(t) = 0$. Then

$$\begin{aligned} E_{X-T | X > T} \left[1 + \frac{\rho}{a}(X - T) \right]^\alpha \left[\ln \left\{ 1 + \frac{\rho}{a}(X - T) \right\} \right]^\beta \\ = \int_0^1 (1 - u)^{-\rho\alpha} [\ln(1 - u)^{-\rho}]^\beta du + O(R_1^*(t^*)), \end{aligned}$$

$$E_{X-T | X > T} \ln \left[1 - \frac{X - T}{Q(1) - T} \right] = \rho + O(R_2^*(t^*)),$$

$$E_{X-T | X > T} \left[1 - \frac{X - T}{Q(1) - T} \right]^\alpha = \frac{1}{-\rho\alpha + 1} + O(R_2^*(t^*)),$$

$$E_{X-T | X > T} \left[\ln \left(1 + \frac{X - T}{T} \right) \right]^\alpha = \int_0^1 [\ln(1 - u)^{-\rho}]^\alpha du + O(R_3^*(t^*)),$$

where $R_1^*(t) = \max\{R(t), R_1(t)\}$, $R_2^*(t) = \max\{R(t), R_2(t)\}$, and $R_3^*(t) = \max\{R(t), R_3(t)\}$.

The proof of this theorem is given in Appendix C.

3. PARAMETER ESTIMATES FROM THE GENERALIZED PARETO DISTRIBUTION

3.1 Method

One approach to estimating ρ and a is to treat the exceedences as a random sample from a parametric model suggested from the conditional distribution of the exceedences. The first of two possible parametric models can be motivated as follows. From (2.3.1),

$$Q_{X-T | X>T}(u; T) = a(t^*)[-g(1-u; -\rho) + \epsilon(t^*, 1-u, \rho)].$$

Taking $a = a(t^*)$ as a scalar given T and $\epsilon(t, u, \rho) = 0$ for all t, u, ρ suggests the *Generalized Pareto Distribution* (GPD) modeling defined below. The GPD model for tail estimates was first proposed by Pickands (1975).

A random variable $W \sim \text{GPD}(\rho, a)$ with $\rho \in \mathbb{R}$, $a > 0$ if it has quantile function

$$Q_{\text{GPD}}(u; \rho, a) = -a \cdot g(1-u; -\rho).$$

Notice that the GPD can also be naturally referred to as the *Power Uniform Distribution* since it can be derived by taking the power transformation of a $\text{Uniform}(0,1)$ random variable.

By inverting the quantile function, the distribution function is

$$F_{\text{GPD}}(w; \rho, a) = \begin{cases} 1 - \left(1 + \frac{\rho w}{a}\right)^{-1/\rho}, & \rho < 0, 0 < w < -a/\rho \\ 1 - e^{-w/a}, & \rho = 0, w > 0 \\ 1 - \left(1 + \frac{\rho w}{a}\right)^{-1/\rho}, & \rho > 0, w > 0, \end{cases}$$

and it follows that the density function is

$$f_{\text{GPD}}(w; \rho, a) = \begin{cases} \frac{1}{a} \left(1 + \frac{\rho w}{a}\right)^{(-1/\rho)-1}, & \rho < 0, 0 < w < -a/\rho \\ \frac{1}{a} e^{-w/a}, & \rho = 0, w > 0 \\ \frac{1}{a} \left(1 + \frac{\rho w}{a}\right)^{(-1/\rho)-1}, & \rho > 0, w > 0. \end{cases}$$

To obtain parameter estimates, choose, as a function of n , a threshold percentile t_n . Then let the threshold value be given by $T_n = Q^-(1 - t_n)$. Compute the exceedences $X_i - Q^-(1 - t_n) = X_i - T_n$ for all $X_i > Q^-(1 - t_n) = T_n$ and treat them as a random sample from a GPD(ρ, a).

3.2 Parameters Estimated By Maximum Likelihood

In other papers using the GPD to estimate tail behavior, maximum likelihood estimation is most popular. For example, DuMouchel (1983), Davison (1984), R. L. Smith (1984, 1987), J. A. Smith (1986), and Joe (1987) propose maximum likelihood to obtain the GPD parameter estimates.

Assuming the exceedences are a random sample from a GPD, the maximum likelihood estimates of ρ and a are the values which maximize the log-likelihood

$$\mathcal{L}_{\text{GPD}}(\rho, a; Y) = \begin{cases} -[nt_n] \ln a \\ \quad - \left(\frac{1}{\rho} + 1 \right) \sum_{i=1}^{[nt_n]} \ln \left(1 + \frac{\rho Y_i}{a} \right), & \rho < 0, \\ \quad \quad \quad a > -\rho \cdot Y([nt_n]; [nt_n]) \\ -[nt_n] \ln a - \frac{1}{a} \sum_{i=1}^{[nt_n]} Y_i, & \rho = 0, a > 0 \\ -[nt_n] \ln a \\ \quad - \left(\frac{1}{\rho} + 1 \right) \sum_{i=1}^{[nt_n]} \ln \left(1 + \frac{\rho Y_i}{a} \right), & \rho > 0, a > 0, \end{cases}$$

where $[\cdot]$ denotes the greatest integer operation and $Y_i = X_i - Q^-(1 - t_n)$ for all $X_i > Q^-(1 - t_n)$, with $Y([nt_n]; [nt_n]) = \max\{Y_1, \dots, Y_{[nt_n]}\}$.

3.2.1 Asymptotic Properties. The asymptotic properties of these estimators do not follow directly from large sample maximum likelihood theory since the exceedences are not a sample from a GPD in general. To derive these results, first express the estimators as solutions to a set of estimating equations, take the Taylor's series expansion, and then compute the asymptotic distribution of each term. This approach yields the following result.

THEOREM 3.2.1. Suppose that $fQ(1-u) = u^{\rho+1}L(u)$, where $-\frac{1}{2} < \rho < \infty$, $\rho \neq 0$ and

$$\left| \frac{L(t)}{L(tu)} - 1 \right| \leq A(u)R(t)$$

for some positive measurable functions $A(u)$ and $R(t)$ where $\lim_{t \rightarrow 0+} R(t) = 0$. Let $\{T_n\}$ be a sequence of threshold values defined on $(Q(0), Q(1))$ such that $nt_n^* \rightarrow \infty$ and $t_n^* \rightarrow 0$ as $n \rightarrow \infty$, where $t_n^* = 1 - F(T_n)$. Further, let $a = a(t_n^*)$ be the scalar value of a function $a(\cdot)$ satisfying $a(t_n^*)hQ(1-t_n^*) = 1 + O(R_1(t_n^*))$ for some positive measurable function $R_1(t)$ with $\lim_{t \rightarrow 0+} R_1(t) = 0$. Let $(\hat{\rho}_n, \hat{a}_n)$ denote the maximum likelihood estimates from the GPD model for the exceedences. Then conditional on T_n ,

$$\begin{bmatrix} \hat{\rho}_n \\ \hat{a}_n \end{bmatrix} \text{ is AN } \left(\begin{bmatrix} \rho + O(R^*(t_n^*)) \\ a + O(R^*(t_n^*)) \end{bmatrix}, ([nt_n^*])^{-1} V_{\text{GPD}} \right)$$

as $nt_n^* \rightarrow \infty$, where $R^*(t) = \max\{R(t), R_1(t)\}$ and

$$V_{\text{GPD}} = \begin{bmatrix} (\rho+1)^2 + O(R^*(t_n^*)) & -a(\rho+1) + O(R^*(t_n^*)) \\ -a(\rho+1) + O(R^*(t_n^*)) & 2a^2(\rho+1) + O(R^*(t_n^*)) \end{bmatrix}.$$

PROOF. Let $(\hat{\rho}_n, \hat{a}_n)$ denote the maximum likelihood estimates derived assuming the exceedences over the threshold T_n are a random sample from a GPD. Then, $(\hat{\rho}_n, \hat{a}_n)$ is the solution to

$$0 = ([nt_n^*])^{-1/2} \sum_{i=1}^{[nt_n^*]} \begin{bmatrix} -\frac{\partial \mathcal{L}_{\text{GPD}}(\hat{\rho}_n, \hat{a}_n; X_i - T_n)}{\partial \rho} \\ -\frac{\partial \mathcal{L}_{\text{GPD}}(\hat{\rho}_n, \hat{a}_n; X_i - T_n)}{\partial a} \end{bmatrix}.$$

Take the first order Taylor's series expansion about the true parameter values (ρ, a) of the right hand side to obtain, for some point (ρ^*, a^*) on the line segment between $(\hat{\rho}_n, \hat{a}_n)$ and (ρ, a) ,

$$0 = ([nt_n^*])^{-1/2} \sum_{i=1}^{[nt_n^*]} \begin{bmatrix} -\frac{\partial \mathcal{L}_{\text{GPD}}(\rho, a; X_i - T_n)}{\partial \rho} \\ -\frac{\partial \mathcal{L}_{\text{GPD}}(\rho, a; X_i - T_n)}{\partial a} \end{bmatrix}$$

$$\begin{aligned}
& + ([nt_n^*])^{-1} \\
& \cdot \sum_{i=1}^{[nt_n^*]} \left[\begin{array}{cc} -\frac{\partial^2 \mathcal{L}_{\text{GPD}}(\rho^*, a^*; X_i - T_n)}{\partial \rho^2} & -\frac{\partial^2 \mathcal{L}_{\text{GPD}}(\rho^*, a^*; X_i - T_n)}{\partial \rho \partial a} \\ -\frac{\partial^2 \mathcal{L}_{\text{GPD}}(\rho^*, a^*; X_i - T_n)}{\partial \rho \partial a} & -\frac{\partial^2 \mathcal{L}_{\text{GPD}}(\rho^*, a^*; X_i - T_n)}{\partial a^2} \end{array} \right] \\
& \cdot ([nt_n^*])^{1/2} \begin{bmatrix} \hat{\rho}_n - \rho \\ \hat{a}_n - a \end{bmatrix}.
\end{aligned}$$

From the Central Limit Theorem for an iid sequence,

$$([nt_n^*])^{-1/2} \sum_{i=1}^{[nt_n^*]} \left[\begin{array}{c} -\frac{\partial \mathcal{L}_{\text{GPD}}(\rho, a; X_i - T_n)}{\partial \rho} \\ -\frac{\partial \mathcal{L}_{\text{GPD}}(\rho, a; X_i - T_n)}{\partial a} \end{array} \right] \text{ is } AN(\mu_n, ([nt_n^*])^{-1} \Sigma_n)$$

as $nt_n^* \rightarrow \infty$, where

$$\mu_n = E \begin{bmatrix} -\frac{\partial \mathcal{L}_{\text{GPD}}(\rho, a; X_i - T_n)}{\partial \rho} \\ -\frac{\partial \mathcal{L}_{\text{GPD}}(\rho, a; X_i - T_n)}{\partial a} \end{bmatrix} \Big| X > T_n = \begin{bmatrix} O(R^*(t_n^*)) \\ O(R^*(t_n^*)) \end{bmatrix},$$

and

$$\begin{aligned}
\Sigma_n &= \text{Cov} \begin{bmatrix} -\frac{\partial \mathcal{L}_{\text{GPD}}(\rho, a; X_i - T_n)}{\partial \rho} \\ -\frac{\partial \mathcal{L}_{\text{GPD}}(\rho, a; X_i - T_n)}{\partial a} \end{bmatrix} \Big| X > T_n \\
&= \begin{bmatrix} \frac{2}{(\rho+1)(2\rho+1)} + O(R^*(t_n^*)) & \frac{1}{a(\rho+1)(2\rho+1)} + O(R^*(t_n^*)) \\ \frac{1}{a(\rho+1)(2\rho+1)} + O(R^*(t_n^*)) & \frac{1}{a^2(2\rho+1)} + O(R^*(t_n^*)) \end{bmatrix}.
\end{aligned}$$

This result follows from the expression of the gradient vector for the GPD log-likelihood which is given in Appendix D and the moments given in Theorem 2.4.1.

For n sufficiently large to make the $O(R^*(t_n^*))$ terms negligible, notice that Σ_n is positive definite if and only if $\rho > -\frac{1}{2}$. The case $\rho < -\frac{1}{2}$ is treated in detail by Smith (1987), but the important change is that the asymptotic distribution of the GPD maximum likelihood estimates is no longer normal, nor is the rate $\sqrt{nt_n^*}$.

If Σ_n is positive definite, then from the Weak Law of Large Numbers and the fact that $(\rho^*, a^*) \xrightarrow{P} (\rho, a)$ as $nt_n^* \rightarrow \infty$, it follows that

$$\Sigma_n^{-1} \cdot ([nt_n^*])^{-1} \sum_{i=1}^{[nt_n^*]} \begin{bmatrix} -\frac{\partial^2 \mathcal{L}_{\text{GPD}}(\rho^*, a^*; X_i - T_n)}{\partial \rho^2} & -\frac{\partial^2 \mathcal{L}_{\text{GPD}}(\rho^*, a^*; X_i - T_n)}{\partial \rho \partial a} \\ -\frac{\partial^2 \mathcal{L}_{\text{GPD}}(\rho^*, a^*; X_i - T_n)}{\partial \rho \partial a} & -\frac{\partial^2 \mathcal{L}_{\text{GPD}}(\rho^*, a^*; X_i - T_n)}{\partial a^2} \end{bmatrix} \xrightarrow{P} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

as $nt_n^* \rightarrow \infty$. This result follows from the expression of the Hessian matrix for the GPD log-likelihood which is given in Appendix D and the moments given in Theorem 2.4.1.

The asymptotic distribution of $(\hat{\rho}_n, \hat{a}_n)$ then follows from Slutsky's Theorem, observing that $V_{\text{GPD}} = \Sigma_n^{-1}$. \square

3.2.2 Computational Aspects of GPD Maximum Likelihood Estimation. This subsection contains a detailed investigation into the problem of maximizing the GPD log-likelihood over the parameter space.

Suppose that Y_1, \dots, Y_k is a random sample from the GPD with largest value

$Y(k; k)$. The log-likelihood is given by

$$\mathcal{L}_{\text{GPD}}(\rho, a; Y) = \begin{cases} -k \ln a - \left(\frac{1}{\rho} + 1\right) \sum_{i=1}^k \ln \left(1 + \frac{\rho Y_i}{a}\right), & \rho < 0, \\ & a > -\rho \cdot Y(k; k) \\ -k \ln a - \frac{1}{a} \sum_{i=1}^k Y_i, & \rho = 0, a > 0 \\ -k \ln a - \left(\frac{1}{\rho} + 1\right) \sum_{i=1}^k \ln \left(1 + \frac{\rho Y_i}{a}\right), & \rho > 0, a > 0. \end{cases}$$

If $\rho < -1$, there is no maximum likelihood estimate since for any $\rho < -1$, $\lim_{a \rightarrow -\rho \cdot Y(k; k)^+} \mathcal{L}_{\text{GPD}}(\rho, a; Y) = \infty$. In order to obtain a finite maximum of the GPD log-likelihood, the constraint $\rho \geq -1$ must be imposed.

There are (in most instances) three values of (ρ, a) which are candidates for the GPD maximum likelihood estimator. The first of these involves the boundary value $\rho = -1$ due to the above constraint on the domain of $\mathcal{L}_{\text{GPD}}(\cdot)$. Given $\rho = -1$, the GPD log-likelihood is maximized at $\hat{a} = Y(k; k)$. This follows since $\mathcal{L}_{\text{GPD}}(\rho = -1, a; Y) = -k \ln a$ is maximized as $a \rightarrow -\rho \cdot Y(k; k)^+ = Y(k; k)^+$. The problem is complicated by the optimization being taken over an open set, but it is treated as a maximum taken over a closed set. Any relative maxima found over the domain of $\mathcal{L}_{\text{GPD}}(\cdot)$ must exceed the GPD log-likelihood evaluated at this boundary in order to be the maximum likelihood estimator.

Figure 2 shows a graph of the GPD log-likelihood function (with a slight modification to permit its definition on the grid required for the graphing routine) for a generated random sample from the GPD. Clearly, there exist relative maxima and minima on the domain of $\mathcal{L}_{\text{GPD}}(\rho, a; Y)$ whose values are found by applying the principles of calculus.

Consider the space defined by $\mathcal{A} = \{-1 < \rho < 0, a > -\rho \cdot Y(k; k)\} \cup \{\rho > 0, a > 0\}$. For some $(\hat{\rho}, \hat{a}) \in \mathcal{A}$, the gradient vector is equal to zero. Using the expression for the gradient vector of the GPD log-likelihood given in Appendix

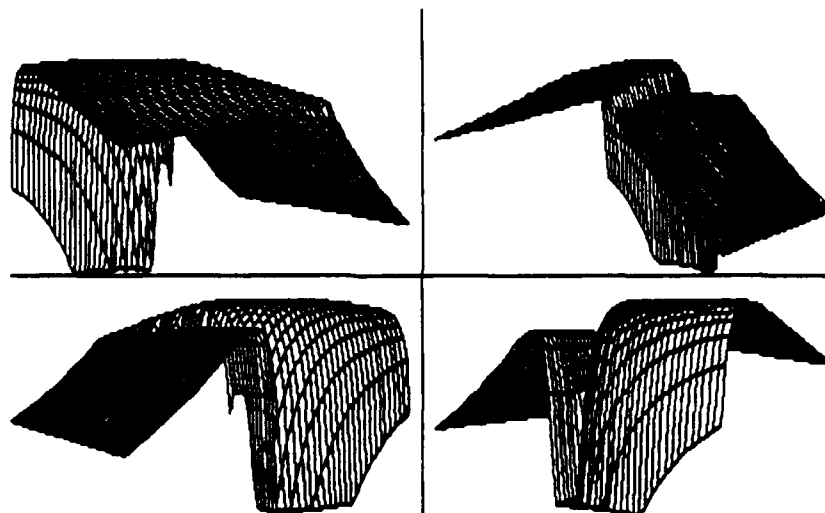


FIG. 2. Graph of the GPD log-likelihood function for a generated random sample. The function has been slightly modified to permit its definition on the grid required for the graphing routine. Notice that there exist relative maxima and minima, implying multiple roots of the gradient vector over the two-dimensional parameter space.

D, the solution to the simultaneous equations may be simplified and written as

$$\left. \begin{aligned} \frac{\partial \mathcal{L}_{\text{GPD}}(\rho, a; Y)}{\partial \rho} &= 0 \\ \frac{\partial \mathcal{L}_{\text{GPD}}(\rho, a; Y)}{\partial a} &= 0 \end{aligned} \right\} \Rightarrow \begin{cases} k(\hat{\rho} + 1) = \sum_{i=1}^k \ln \left(1 + \frac{\hat{\rho} Y_i}{\hat{a}} \right) \\ \quad + (\hat{\rho} + 1) \sum_{i=1}^k \left(1 + \frac{\hat{\rho} Y_i}{\hat{a}} \right)^{-1} \\ k = (\hat{\rho} + 1) \sum_{i=1}^k \left(1 + \frac{\hat{\rho} Y_i}{\hat{a}} \right)^{-1} \end{cases}$$

$$\Rightarrow \begin{cases} \left[1 + \frac{1}{k} \sum_{i=1}^k \ln \left(1 + \frac{\hat{\rho} Y_i}{\hat{a}} \right) \right] \cdot \left[\frac{1}{k} \sum_{i=1}^k \left(1 + \frac{\hat{\rho} Y_i}{\hat{a}} \right)^{-1} \right] = 1 \\ \hat{\rho} = \frac{1}{k} \sum_{i=1}^k \ln \left(1 + \frac{\hat{\rho} Y_i}{\hat{a}} \right). \end{cases}$$

The bivariate search for the zeroes of the gradient vector over \mathcal{A} can be reduced to a univariate search since the second equation is a closed form representation for the estimator of ρ given the ratio $\hat{\rho}/\hat{a}$, and the first equation depends only on $\hat{\rho}/\hat{a}$. Therefore, the zeroes of the gradient of the log-likelihood of the GPD are the solution(s) to $h(\hat{\rho}/\hat{a}) = 0$, where the function $h(\cdot)$ is defined by

$$h(\eta) = \left[1 + \frac{1}{k} \sum_{i=1}^k \ln(1 + \eta Y_i) \right] \cdot \left[\frac{1}{k} \sum_{i=1}^k (1 + \eta Y_i)^{-1} \right] - 1,$$

with domain $\{\eta > -1/Y(k; k), \eta \neq 0\}$.

An example of the function $h(\eta)$ is given in Figure 3 for the random sample from the GPD used in Figure 2. Notice that there are two roots of $h(\cdot)$ in this example, only one of which corresponds to the local maximum. It is easily shown that $h(\cdot)$ is continuous at zero since $\lim_{\eta \rightarrow 0} h(\eta) = 0$.

A more important consequence of this limit is that $\eta \rightarrow 0$ gives the second possible value for the GPD maximum likelihood estimate. The limit $\eta \rightarrow 0$ corresponds to the case $\rho = 0$, where $\mathcal{L}_{\text{GPD}}(\cdot)$ is only a function of a . The extremum $\hat{\rho} = 0$, $\hat{a} = 1/\bar{Y}$, where $\bar{Y} = k^{-1} \sum_{i=1}^k Y_i$, follows from solving

$$\frac{\partial \mathcal{L}_{\text{GPD}}(\rho = 0, a; Y)}{\partial a} = \frac{1}{a^2} \sum_{i=1}^k Y_i - \frac{k}{a} = 0$$

which is a local maximum if

$$\frac{\partial^2 \mathcal{L}_{\text{GPD}}(\hat{\rho} = 0, \hat{a}; Y)}{\partial a^2} = \frac{k}{\hat{a}^2} - \frac{2}{\hat{a}^3} \sum_{i=1}^k Y_i < 0 \iff \hat{a} < \sqrt{2}.$$

The third possible value for the maximum likelihood estimate which exists in most cases must be found numerically. Subsection 3.2.3 describes an algorithm which determines if such a root exists and, assuming it does exist, finds it using

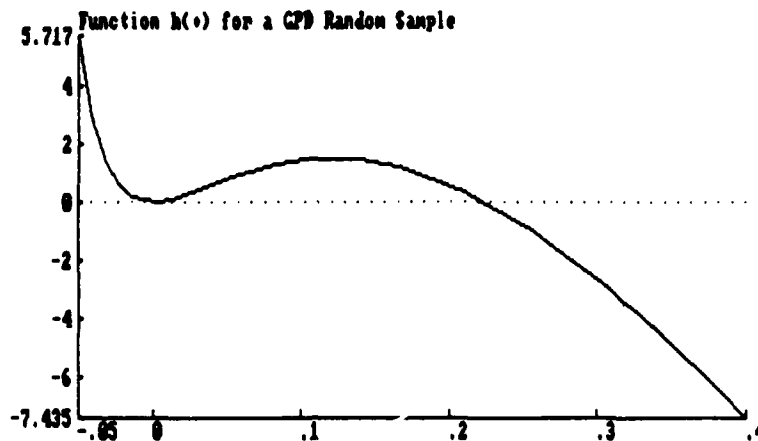


FIG. 3. Graph of the function $h(\eta)$ used to simplify the search for the zeroes of the gradient vector on the two-dimensional parameter space for a GPD random sample. Notice that there are two zeroes of the function, only one of which corresponds to the local maximum.

the bisection root finding algorithm. If such a root is found, the third possible value for the GPD maximum likelihood estimate can be computed from

$$\hat{\rho} = \frac{1}{k} \sum_{i=1}^k \ln(1 + \hat{\eta} Y_i)$$

$$\hat{a} = \frac{\hat{\rho}}{\hat{\eta}}.$$

This relative extremum must be verified to be a local maximum by considering the Hessian matrix of the GPD log-likelihood given in Appendix D. The point $(\hat{\rho}, \hat{a})$ is a local maximum if the Hessian matrix evaluated at the estimators is negative definite.

3.2.3 Proposed Algorithm For The GPD Maximum Likelihood Estimator. Hosking and Wallis (1987) attempted to find the GPD maximum likelihood estimator using Newton-Raphson optimization in two dimensions and found that their algorithm failed to converge to a local maximum with alarming frequency. The table of failures to converge given by Hosking and Wallis (1987) is reproduced in Table 2.

TABLE 2

Failure rate of the Newton-Raphson optimization in two dimensions of the log-likelihood of the GPD, reproduced from Hosking and Wallis (1987). Tabulated values are the number of failures to converge of Newton-Raphson per 100 simulated samples. The GPD log-likelihood has a maximum, but this algorithm fails far too frequently to be considered reliable in practice.

	$\rho = -.4$	$\rho = -.2$	$\rho = 0$	$\rho = .2$	$\rho = .4$
$n = 15$	41.7	22.7	12.2	4.8	3.6
$n = 25$	14.6	4.7	1.5	.3	.2

The observed failure to converge of the Newton-Raphson algorithm has two explanations. First, the root of the gradient vector for the GPD log-likelihood may not satisfy the second order conditions to be a local maximum. The GPD log-likelihood often has two zeroes of the gradient vector, so the root obtained by the Newton-Raphson algorithm using a particular initial value may not be the local maximum.

Second, in terms of the parameters (ρ, a) , the root of the gradient vector corresponding to $\eta \rightarrow 0$ is $\rho = 0$. However, it is interesting to note that a numerical optimization routine in two dimensions may increase a at each iteration in an attempt to find the zero of the gradient vector. Such behavior is due to the reparameterization $\eta = \rho/a$, where η can be made arbitrarily close to zero by letting $a \rightarrow \infty$. Therefore, the gradient vector may always be made closer to zero with a larger value of a . The Newton-Raphson algorithm in two dimensions can continually increase a until reaching an upper bound for the iterations which then signals that the algorithm failed to converge.

An algorithm which computes each of the three possible values for the GPD maximum likelihood estimator discussed in Subsection 3.2.2 and then selects the maxima of the GPD log-likelihood is given by:

- Choose a δ such that for $|\hat{\eta}| < \delta$, it will be considered that $\hat{\eta} = 0$ and there is a lone solution to $h(\eta) = 0$. For example, let $\delta = .0001$.
- Choose an ϵ to be used as a convergence criterion such that for $|h(\hat{\eta})| < \epsilon$, it will be considered that $\hat{\eta}$ satisfies $h(\hat{\eta}) = 0$. For example, let $\epsilon = .0001$.

- Compute $h(-\delta)$ and $h(\delta)$.
- If $h(-\delta) < 0$, then:
 1. Let $\eta^{(L)} = [-1/Y(k; k)] \cdot [i/(i+1)]$ for the smallest $i \in \{1, 2, \dots, M\}$ such that $h(\eta^{(L)}) > 0$, where M is a specified bound on the number of iterations. If there is no $\eta^{(L)}$ such that $h(\eta^{(L)}) > 0$, then a nonzero root cannot be found numerically.
 2. Let $\eta^{(U)} = -\delta$.
 3. Use the bisection algorithm to find $\hat{\eta}$ such that $|h(\hat{\eta})| < \epsilon$.
- If $h(\delta) > 0$, then:
 1. Let $\eta^{(L)} = \delta$.
 2. Let $\eta^{(U)} = i$ for the smallest $i \in \{1, 2, \dots, M\}$ such that $h(\eta^{(U)}) < 0$, where M is a specified bound on the number of iterations. If there is no $\eta^{(U)}$ such that $h(\eta^{(U)}) < 0$, then a nonzero root cannot be found numerically.
 3. Use the bisection algorithm to find $\hat{\eta}$ such that $|h(\hat{\eta})| < \epsilon$.
- If a nonzero $\hat{\eta}$ is found such that $|h(\hat{\eta})| < \epsilon$, then:
 1. Compute $\hat{\rho}_1 = (1/k) \sum_{i=1}^k \ln(1 + \hat{\eta} Y_i)$ and $\hat{a}_1 = \hat{\rho}_1 / \hat{\eta}$.
 2. Compute the Hessian at $(\hat{\rho}_1, \hat{a}_1)$.
 3. If the Hessian at $(\hat{\rho}_1, \hat{a}_1)$ is negative definite, then $(\hat{\rho}_1, \hat{a}_1)$ is a local maximum of the GPD log-likelihood.
- If $\bar{Y} > 1/\sqrt{2}$, then $\hat{\rho}_2 = 0$ and $\hat{a}_2 = 1/\bar{Y}$ is also a local maximum of the GPD log-likelihood.
- The relative maximum is $(\hat{\rho}_1, \hat{a}_1)$ if $\mathcal{L}_{\text{GPD}}(\hat{\rho}_1, \hat{a}_1; Y) > \mathcal{L}_{\text{GPD}}(\hat{\rho}_2, \hat{a}_2; Y)$. Otherwise, the relative maximum is $(\hat{\rho}_2, \hat{a}_2)$.
- If $\mathcal{L}_{\text{GPD}}(\hat{\rho}, \hat{a}; Y) > -k \ln Y(k; k)$ where $(\hat{\rho}, \hat{a})$ denote the relative maximum, then $\hat{\rho}$ and \hat{a} are the GPD maximum likelihood estimates. Otherwise, the boundary maxima $\hat{\rho} = -1$ and $\hat{a} = Y(k; k)$ are the GPD maximum likelihood estimates.

It should be noted that the bisection algorithm is preferred for the numerical root finding because it ensures that the nonzero root will be found if it exists.

The bisection algorithm assumes two values $\eta^{(L)}$ and $\eta^{(U)}$ are given such that $h(\eta^{(L)}) \cdot h(\eta^{(U)}) < 0$, and a convergence criterion ϵ (which was defined earlier) is specified. Given these values, the *bisection algorithm* is defined as:

- Compute $\hat{\eta} = (\eta^{(L)} + \eta^{(U)})/2$.
- If $|h(\hat{\eta})| \leq \epsilon$ then terminate the algorithm, returning $\hat{\eta}$.
- If $|h(\hat{\eta})| > \epsilon$ then
 1. If $h(\hat{\eta}) \cdot h(\eta^{(L)}) < 0$ then set $\eta^{(U)} = \hat{\eta}$ and repeat the algorithm.
 2. If $h(\hat{\eta}) \cdot h(\eta^{(U)}) < 0$ then set $\eta^{(L)} = \hat{\eta}$ and repeat the algorithm.

Typically, the bisection algorithm is criticized for its slow rate of convergence. However, in the application of this algorithm in simulation studies, the nonzero root of $h(\cdot)$ is most often found in five to ten iterations.

3.3 Parameters Estimated From The Sample Quantile Process

Other parameter estimates can be obtained by applying the theory of regression analysis on continuous-parameter time series from the reproducing kernel Hilbert space (RKHS) point of view given by Parzen (1961a,b, 1967) applied to the sample quantile process $Q^{\sim}(u)$. This approach follows the ideas in Parzen (1979) for estimating location and scale parameters.

Parzen (1979) motivates this approach by first stating the following theorem on the strong approximation of the quantile process.

THEOREM [Csörgő and Révész (1978)]. Let $\{Q^{\sim}(u), 0 < u < 1\}$ denote the sample quantile process of a random sample from a population with continuous distribution function $F_0(x)$, quantile function $Q_0(u)$, density function $f_0(x)$, and density-quantile function $f_0 Q_0(1 - u) = u^{\rho+1} L(u)$ for $\rho \in \mathbb{R}$ and $L(u)$ slowly varying as $u \rightarrow 0^+$. Let $\{Q_U^{\sim}(u), 0 < u < 1\}$ denote the quantile process of the uniformly distributed random variables $U_i = F(X_i)$. Let

$$R_n = \sup_{0 < u < 1} \sqrt{n} |f_0 Q_0(u) \{Q^{\sim}(u) - Q_0(u)\} - \{Q_U^{\sim}(u) - u\}|.$$

Then almost surely,

$$R_n = \begin{cases} O(n^{-1/2} \ln \ln n), & \text{if } \rho < 0 \\ O(n^{-1/2} (\ln \ln n)^2), & \text{if } \rho = 0 \\ O(n^{-1/2} (\ln \ln n)^{\rho+1} (\ln n)^{\rho(1+\epsilon)}), & \text{if } \rho > 0 \end{cases}$$

for every $\epsilon > 0$.

Therefore, $f_0 Q_0(u) \{Q^-(u) - Q_0(u)\}$ can be approximated by the uniform sample quantile process $\{Q_U^-(u), 0 < u < 1\}$ whose weak convergence is considered in the following theorem. Define a Brownian bridge to be a zero mean normal process with covariance kernel $K_B(s, t) = \min\{s, t\} - st$, $0 \leq s, t \leq 1$.

THEOREM [Csörgő and Révész (1975)]. A Brownian bridge $\{B_n(u), 0 \leq u \leq 1\}$ can be defined for each n such that, almost surely,

$$\sup_{0 \leq u \leq 1} |\sqrt{n} \{Q_U^-(u) - u\} - B_n(u)| = O(n^{-1/2} \ln n).$$

These two theorems are then interpreted for purposes of statistical inference to mean that $\sqrt{n} f_0 Q_0(u) [Q^-(u) - Q_0(u)]$ is distributed as a Brownian bridge $B(u)$.

The GPD model for exceedences of a threshold assumes that for given t_n ,

$$Q_{X-T_n} | X > Q^-(1-t_n)=T_n (u; t_n, T_n) = -a \cdot g(1-u; -\rho)$$

with

$$f Q_{X-T_n} | X > Q^-(1-t_n)=T_n (u; t_n, T_n) = \frac{(1-u)^{\rho+1}}{a}.$$

The sample quantile process for the exceedences for given t_n is

$$Q_{X-T_n} | X > Q^-(1-t_n)=T_n (u; t_n, T_n) = Q^-(1-t_n(1-u)) - Q^-(1-t_n).$$

Therefore, estimating ρ and a becomes a problem in regression analysis of continuous-parameter time series by writing

$$\begin{aligned} (1-u)^{\rho+1} [Q^-(1-t_n(1-u)) - Q^-(1-t_n)] \\ = a(1-u)^{\rho+1} g(1-u; -\rho) + \sigma_B B(u) \end{aligned}$$

where $\sigma_B = a/\sqrt{[nt_n]}$ is treated as a free parameter not constrained to be related to a .

Estimators can be formed from this time series regression analysis after a reproducing kernel inner product is found corresponding to the Brownian bridge covariance kernel $K_B(s, t)$. This RKHS consists of L_2 differentiable functions f, g on the interval $p \leq u \leq q$ with inner product

$$\langle f, g \rangle_{p,q} = \int_p^q f'(u)g'(u)du + \frac{1}{p}f(p)g(p) + \frac{1}{1-q}f(q)g(q).$$

Parzen (1979) proves the reproducing formula $\langle f, K_B(\cdot, t) \rangle_{p,q} = f(t)$ for $p \leq t \leq q$ which verifies $\langle f, g \rangle_{p,q}$ is the reproducing kernel inner product.

Applying the ideas of modeling the quantile process as a continuous parameter time series, Parzen (1979) derives optimal estimates (along with the corresponding influence functions) for location and scale parameters. This approach does not meet with the same success in the GPD modeling for tail estimates since the quantile process is expressed as a nonlinear function of the parameters. However, given ρ the model is a linear function in a .

Drawing from the applications paradigm of the Box-Cox (1964) transformation in regression, estimates of (ρ, a) are found by:

1. Choose a reasonable range of values for ρ .
2. For each value $\rho^{(i)}$, compute

$$\hat{a}_{p,q}^{(i)} = \frac{\left\langle (1-u)^{\rho+1}g(1-u; -\rho), (1-u)^{\rho+1}[Q^-(1-t_n(1-u)) - Q^-(1-t_n)] \right\rangle_{p,q}}{\left\langle (1-u)^{\rho+1}g(1-u; -\rho), (1-u)^{\rho+1}g(1-u; -\rho) \right\rangle_{p,q}},$$

and compute the value of a specified loss function $R(\rho^{(i)}, \hat{a}_{p,q}^{(i)})$.

3. Choose the estimate of (ρ, a) which minimizes $R(\cdot)$.

4. PARAMETER ESTIMATES FROM THE GENERALIZED EXTREME VALUE DISTRIBUTION

4.1 Method

A second parametric model for the conditional distribution of the exceedences can be motivated as follows. From (2.3.1),

$$Q_{X-T | X>T}(u; T) = a(t^*)[-g(1-u; -\rho) + \epsilon(t^*, 1-u, \rho)].$$

Taking $a = a(t^*)$ as a scalar given T , $\epsilon(t, u, \rho) = 0$ for all t, u, ρ , and noticing that as $u \rightarrow 1^-$, $g(-\ln u; -\rho) \sim g(1-u; -\rho)$ suggests the *Generalized Extreme Value Distribution* (GEV) defined below. The GEV probability model is used extensively in practical applications involving floods and extreme sea levels where the random variable is the maximum value over a given time period. Jenkinson (1955), the Flood Studies Report [Natural Environment Research Council (NERC) (1975)], and Blackman and Graff (1978) give examples of GEV applications.

In this work, it is important to point out that the GEV is used as a model for the exceedences. This alters both the application of the model and the properties of the estimates from the usual scenarios where the entire sample is modeled as GEV.

A random variable $W \sim \text{GEV}(\rho, a)$ with $\rho \in \mathbb{R}$, $a > 0$ if it has quantile function

$$Q_{\text{GEV}}(u; \rho, a) = -a \cdot g(-\ln u; -\rho).$$

The name Generalized Extreme Value distribution follows from its unifying representation of the three types of extreme value distributions derived by Fisher and Tippet (1928). However, the GEV can also be naturally referred to as the *Power Exponential Distribution* since it can be derived by taking the power transformation of an Exponential(1) random variable.

By inverting the quantile function, the distribution function is

$$F_{\text{GEV}}(w; \rho, a) = \begin{cases} \exp \left\{ - \left(1 + \frac{\rho w}{a} \right)^{-1/\rho} \right\}, & \rho < 0, 0 < w < -a/\rho \\ \exp \{ -e^{-w/a} \}, & \rho = 0, w > 0 \\ \exp \left\{ - \left(1 + \frac{\rho w}{a} \right)^{-1/\rho} \right\}, & \rho > 0, w > 0, \end{cases}$$

and it follows that the density function is

$$f_{\text{GEV}}(w; \rho, a) = \begin{cases} \frac{1}{a} \left(1 + \frac{\rho w}{a} \right)^{(-1/\rho)-1} \\ \quad \cdot \exp \left\{ - \left(1 + \frac{\rho w}{a} \right)^{-1/\rho} \right\}, & \rho < 0, 0 < w < -a/\rho \\ \frac{1}{a} e^{-w/a} \\ \quad \cdot \exp \{ -e^{-w/a} \}, & \rho = 0, w > 0 \\ \frac{1}{a} \left(1 + \frac{\rho w}{a} \right)^{(-1/\rho)-1} \\ \quad \cdot \exp \left\{ - \left(1 + \frac{\rho w}{a} \right)^{-1/\rho} \right\}, & \rho > 0, w > 0. \end{cases}$$

To obtain tail estimates, choose, as a function of n , a threshold percentile t_n . Then let the threshold value be given by $T_n = Q^*(1 - t_n)$. Compute the exceedences $X_i - T_n = X_i - Q^*(1 - t_n)$ for all $X_i > T_n = Q^*(1 - t_n)$ and treat them as a random sample from a $\text{GEV}(\rho, a)$.

4.2 Parameters Estimated By Maximum Likelihood

Assuming the exceedences are a random sample from a GEV distribution, the maximum likelihood estimates of ρ and a are the values which maximize the

log-likelihood

$$\mathcal{L}_{\text{GEV}}(\rho, a; Y) = \begin{cases} \begin{aligned} & -[nt_n] \ln a \\ & - \left(\frac{1}{\rho} + 1 \right) \sum_{i=1}^{[nt_n]} \ln \left(1 + \frac{\rho Y_i}{a} \right) \\ & - \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a} \right)^{-1/\rho}, \end{aligned} & \begin{aligned} & \rho < 0, \\ & a > -\rho \cdot Y([nt_n]; [nt_n]) \end{aligned} \\ \\ \begin{aligned} & -[nt_n] \ln a - \frac{1}{a} \sum_{i=1}^{[nt_n]} Y_i \\ & - \sum_{i=1}^{[nt_n]} e^{-Y_i/a}, \end{aligned} & \rho = 0, a > 0 \\ \\ \begin{aligned} & -[nt_n] \ln a \\ & - \left(\frac{1}{\rho} + 1 \right) \sum_{i=1}^{[nt_n]} \ln \left(1 + \frac{\rho Y_i}{a} \right) \\ & - \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a} \right)^{-1/\rho}, \end{aligned} & \rho > 0, a > 0, \end{cases}$$

where $[\cdot]$ denotes the greatest integer operation and $Y_i = X - Q^-(1 - t_n)$ for all $X_i > Q^-(1 - t_n)$, with $Y([nt_n]; [nt_n]) = \max\{Y_1, \dots, Y_{[nt_n]}\}$.

Maximum likelihood estimation for the GEV is often criticized because it must be performed numerically. However, Hosking (1985) provides an algorithm for finding the GEV maximum likelihood estimates for a random sample based on Newton-Raphson iteration with some modifications designed to improve the rate of convergence. This algorithm performs well for $|\rho| < \frac{1}{2}$ and $nt_n > 15$.

The large sample properties of these estimators do not follow from directly from large sample maximum likelihood theory since the exceedences are not a sample from a GEV distribution in general. To derive the asymptotic results, first express the estimators as solutions to a set of estimating equations, take the

Taylor's series expansion, and then compute the asymptotic distribution of each term. This approach yields the following result.

THEOREM 4.2.1. *Suppose that $fQ(1-u) = u^{\rho+1}L(u)$, where $\rho_0 < \rho < 1$, $\rho \neq 0$, where ρ_0 is the only real root of $h(\rho)$ on the interval $(-1, 1)$ where*

$$(4.2.1) \quad h(\rho) = 24\rho^{17} - 46\rho^{16} - 345\rho^{15} + 520\rho^{14} + 1715\rho^{13} \\ - 1490\rho^{12} - 3877\rho^{11} + 584\rho^{10} + 5729\rho^9 - 6428\rho^8 \\ - 7150\rho^7 + 24532\rho^6 - 2184\rho^5 - 2672\rho^4 - 7072\rho^3 \\ - 8512\rho^2 - 4224\rho - 768.$$

Approximately, $\rho_0 \approx -0.356967$. Also, suppose

$$\left| \frac{L(t)}{L(tu)} - 1 \right| \leq A(u)R(t)$$

for some positive measurable functions $A(u)$ and $R(t)$ where $\lim_{t \rightarrow 0+} R(t) = 0$. Let $\{T_n\}$ be a sequence of thresholds defined on $(Q(0), Q(1))$ such that $nt_n^* \rightarrow \infty$ and $t_n^* \rightarrow 0$ as $n \rightarrow \infty$, where $t_n^* = 1 - F(T_n)$. Further, let $a = a(t_n^*)$ be the scalar value of a function $a(\cdot)$ satisfying $a(t_n^*)hQ(1-t_n^*) = 1 + O(R_1(t_n^*))$ for some positive measurable function $R_1(t)$ with $\lim_{t \rightarrow 0+} R_1(t) = 0$. Let $(\hat{\rho}_n, \hat{a}_n)$ denote the maximum likelihood estimates from the GEV model for the exceedences. Then conditional on T_n ,

$$\begin{bmatrix} \hat{\rho}_n \\ \hat{a}_n \end{bmatrix} \text{ is AN } \left(\begin{bmatrix} \rho - \frac{\rho-4}{4\rho(\rho-2)^2} + O(R^*(t_n^*)) \\ a + \frac{1}{2a(\rho+2)} + O(R^*(t_n^*)) \end{bmatrix}, ([nt_n^*])^{-1}V_{\text{GEV}} \right)$$

as $nt_n^* \rightarrow \infty$, where $R^*(t) = \max\{R(t), R_1(t)\}$ and

$$V_{\text{GEV}} = \frac{1}{h(\rho)} \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix},$$

with $h(\rho)$ given in (4.2.1) and

$$v_{11} = 4\rho^4(\rho-2)^4(\rho-1)^3(\rho+1)^2(\rho+2)^3(2\rho+1)(3\rho^2+4\rho+2) + O(R^*(t_n^*)),$$

$$\begin{aligned}
v_{22} = & 2a^2\rho^2(\rho-2)(\rho+1)(\rho+2)^3(2\rho+1) \\
& \cdot (8\rho^{11} - 50\rho^{10} + 81\rho^9 + 63\rho^8 - 374\rho^7 + 668\rho^6 \\
& - 623\rho^5 + 203\rho^4 - 16\rho^3 - 32\rho^2 + 96\rho + 48) + O(R^*(t_n^*)),
\end{aligned}$$

$$\begin{aligned}
v_{12} = & v_{21} \\
= & 4a\rho^5(\rho-2)^2(\rho-1)^3(\rho+1)(\rho+2)^2(2\rho+1) \\
& \cdot (\rho^4 + 6\rho^3 - 21\rho^2 - 48\rho + 16) + O(R^*(t_n^*)).
\end{aligned}$$

PROOF. Let $(\hat{\rho}_n, \hat{a}_n)$ denote the maximum likelihood estimates derived assuming the exceedences over the threshold T_n are a random sample from a GEV distribution. Then, $(\hat{\rho}_n, \hat{a}_n)$ is the solution to

$$0 = ([nt_n^*])^{-1/2} \sum_{i=1}^{[nt_n^*]} \begin{bmatrix} -\frac{\partial \mathcal{L}_{\text{GEV}}(\hat{\rho}_n, \hat{a}_n; X_i - T_n)}{\partial \rho} \\ -\frac{\partial \mathcal{L}_{\text{GEV}}(\hat{\rho}_n, \hat{a}_n; X_i - T_n)}{\partial a} \end{bmatrix}.$$

Take the first order Taylor's series expansion about the true parameter values (ρ, a) of the right hand side to obtain, for some point (ρ^*, a^*) on the line segment between $(\hat{\rho}_n, \hat{a}_n)$ and (ρ, a) ,

$$\begin{aligned}
0 = & ([nt_n^*])^{-1/2} \sum_{i=1}^{[nt_n^*]} \begin{bmatrix} -\frac{\partial \mathcal{L}_{\text{GEV}}(\rho, a; X_i - T_n)}{\partial \rho} \\ -\frac{\partial \mathcal{L}_{\text{GEV}}(\rho, a; X_i - T_n)}{\partial a} \end{bmatrix} \\
& + ([nt_n^*])^{-1} \sum_{i=1}^{[nt_n^*]} \begin{bmatrix} -\frac{\partial^2 \mathcal{L}_{\text{GEV}}(\rho^*, a^*; X_i - T_n)}{\partial \rho^2} - \frac{\partial^2 \mathcal{L}_{\text{GEV}}(\rho^*, a^*; X_i - T_n)}{\partial \rho \partial a} \\ -\frac{\partial^2 \mathcal{L}_{\text{GEV}}(\rho^*, a^*; X_i - T_n)}{\partial \rho \partial a} - \frac{\partial^2 \mathcal{L}_{\text{GEV}}(\rho^*, a^*; X_i - T_n)}{\partial a^2} \end{bmatrix}
\end{aligned}$$

$$([nt_n^*])^{1/2} \begin{bmatrix} \hat{\rho}_n - \rho \\ \hat{a}_n - a \end{bmatrix}.$$

From the Central Limit Theorem for an iid sequence,

$$([nt_n^*])^{-1/2} \sum_{i=1}^{[nt_n^*]} \begin{bmatrix} -\frac{\partial \mathcal{L}_{\text{GEV}}(\rho, a; X_i - T_n)}{\partial \rho} \\ -\frac{\partial \mathcal{L}_{\text{GEV}}(\rho, a; X_i - T_n)}{\partial a} \end{bmatrix} \text{ is } AN(\mu_n, ([nt_n^*])^{-1} \Sigma_n)$$

as $nt_n^* \rightarrow \infty$, where

$$\begin{aligned} \mu_n &= E_{X-T_n | X > T_n} \begin{bmatrix} -\frac{\partial \mathcal{L}_{\text{GEV}}(\rho, a; X_i - T_n)}{\partial \rho} \\ -\frac{\partial \mathcal{L}_{\text{GEV}}(\rho, a; X_i - T_n)}{\partial a} \end{bmatrix} \\ &= \begin{bmatrix} -\frac{\rho - 4}{4\rho(\rho - 2)^2} + O(R^*(t_n^*)) \\ \frac{1}{2a(\rho + 2)} + O(R^*(t_n^*)) \end{bmatrix}, \end{aligned}$$

and

$$\begin{aligned} \Sigma_n &= \text{Cov}_{X-T_n | X > T_n} \left(\begin{bmatrix} -\frac{\partial \mathcal{L}_{\text{GEV}}(\rho, a; X_i - T_n)}{\partial \rho} \\ -\frac{\partial \mathcal{L}_{\text{GEV}}(\rho, a; X_i - T_n)}{\partial a} \end{bmatrix} \begin{bmatrix} -\frac{\partial \mathcal{L}_{\text{GEV}}(\rho, a; X_i - T_n)}{\partial \rho} \\ -\frac{\partial \mathcal{L}_{\text{GEV}}(\rho, a; X_i - T_n)}{\partial a} \end{bmatrix} \right) \\ &= \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}, \end{aligned}$$

with

$$\sigma_{11} = \left[\frac{1}{4\rho^4(\rho - 2)^3(\rho - 1)^3(\rho + 1)(\rho + 2)(2\rho + 1)} \right]$$

$$\begin{aligned}
& \cdot (8\rho^{11} - 50\rho^{10} + 81\rho^9 + 63\rho^8 - 374\rho^7 + 668\rho^6 \\
& \quad - 623\rho^5 + 203\rho^4 - 16\rho^3 - 32\rho^2 + 96\rho + 48) + O(R^*(t_n^*)), \\
\sigma_{22} &= \frac{3\rho^2 + 4\rho + 2}{2a^2\rho^2(\rho + 2)(2\rho + 1)} + O(R^*(t_n^*)), \\
\sigma_{12} &= \sigma_{21} \\
&= -\frac{(\rho^4 + 6\rho^3 - 21\rho^2 - 48\rho + 16)}{2a\rho(\rho - 2)^2(\rho + 1)(\rho + 2)^2(2\rho + 1)} + O(R^*(t_n^*)).
\end{aligned}$$

This result follows from the expression of the gradient vector for the GEV log-likelihood which is given in Appendix E and the moments given in Theorem 2.4.1.

For n sufficiently large to make the $O(R^*(t_n^*))$ terms negligible, notice that Σ_n is positive definite if and only if $\rho_0 < \rho < 1$, where ρ_0 is the only real root of $h(\rho)$ on the interval $(-1, 1)$ for $h(\rho)$ given in (4.2.1). Approximately, $\rho_0 \approx -0.356967$.

If Σ_n is positive definite, then from the Weak Law of Large Numbers and the fact that $(\rho^*, a^*) \xrightarrow{p} (\rho, a)$ as $nt_n^* \rightarrow \infty$, it follows that

$$\begin{aligned}
& \Sigma_n^{-1} \cdot ([nt_n^*])^{-1} \sum_{i=1}^{[nt_n^*]} \left[\begin{array}{cc} -\frac{\partial^2 \mathcal{L}_{\text{GEV}}(\rho^*, a^*; X_i - T_n)}{\partial \rho^2} & -\frac{\partial^2 \mathcal{L}_{\text{GEV}}(\rho^*, a^*; X_i - T_n)}{\partial \rho \partial a} \\ -\frac{\partial^2 \mathcal{L}_{\text{GEV}}(\rho^*, a^*; X_i - T_n)}{\partial \rho \partial a} & -\frac{\partial^2 \mathcal{L}_{\text{GEV}}(\rho^*, a^*; X_i - T_n)}{\partial a^2} \end{array} \right] \\
& \xrightarrow{p} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
\end{aligned}$$

as $nt_n^* \rightarrow \infty$. This result follows from the expression of the Hessian matrix for the GEV log-likelihood which is given in Appendix E and the moments given in Theorem 2.4.1.

The asymptotic distribution of $(\hat{\rho}_n, \hat{a}_n)$ then follows from Slutsky's Theorem, observing that $V_{\text{GEV}} = \Sigma_n^{-1}$. \square

4.3 Parameters Estimated From The Sample Quantile Process

Other parameter estimates can be obtained by applying the theory of regression analysis on continuous-parameter time series from the reproducing kernel Hilbert space (RKHS) point of view to the sample quantile process $Q^{\sim}(u)$. The justification for this approach was given in Subsection 3.3, but the main consequence to statistical inference is that for a random sample from a population with quantile function $Q_0(u)$ and density-quantile function $f_0 Q_0(u)$,

$$\sqrt{n} f_0 Q_0(u) [Q^{\sim}(u) - Q_0(u)] \sim B(u)$$

where $B(u)$ is a Brownian bridge.

The GEV model for exceedences of a threshold assumes that for given t_n ,

$$Q_{X-T_n} | X > Q^{\sim}(1-t_n)=T_n (u; t_n, T_n) = -a \cdot g(-\ln u; -\rho)$$

with

$$f Q_{X-T_n} | X > Q^{\sim}(1-t_n)=T_n (u; t_n, T_n) = \frac{u(-\ln u)^{\rho+1}}{a}.$$

The sample quantile process for the exceedences for given t_n is

$$Q^{\sim}_{X-T_n} | X > Q^{\sim}(1-t_n)=T_n (u; t_n, T_n) = Q^{\sim}(1 - t_n(1 - u)) - Q^{\sim}(1 - t_n).$$

Therefore, estimating ρ and a becomes a problem in regression analysis of continuous-parameter time series by writing

$$\begin{aligned} u(-\ln u)^{\rho+1} [Q^{\sim}(1 - t_n(1 - u)) - Q^{\sim}(1 - t_n)] \\ = a u(-\ln u)^{\rho+1} g(-\ln u; -\rho) + \sigma_B B(u) \end{aligned}$$

where $\sigma_B = a/\sqrt{[nt_n]}$ is treated as a free parameter not constrained to be related to a .

Estimators can be formed from this time series regression using the reproducing kernel inner product corresponding to the Brownian bridge covariance kernel $K_B(s, t)$. This RKHS consists of L_2 differentiable functions f, g on the interval $p \leq u \leq q$ with inner product

$$\langle f, g \rangle_{p,q} = \int_p^q f'(u)g'(u)du + \frac{1}{p}f(p)g(p) + \frac{1}{1-q}f(q)g(q).$$

Since this model is not linear in the parameters, no closed form expressions exist for the estimators. However, drawing from the applications paradigm of the Box-Cox (1964) transformation in regression, estimates of (ρ, a) are found from $Q^*(\cdot)$ by:

1. Choose a reasonable range of values for ρ .
2. For each value of $\rho^{(i)}$ the model is linear in a , so compute

$$\hat{a}_{p,q}^{(i)} = \frac{\left\langle u(-\ln u)^{\rho+1}g(-\ln u; -\rho), u(-\ln u)^{\rho+1}[Q^*(1 - t_n(1 - u)) - Q^*(1 - t_n)] \right\rangle_{p,q}}{\left\langle u(-\ln u)^{\rho+1}g(-\ln u; -\rho), u(-\ln u)^{\rho+1}g(-\ln u; -\rho) \right\rangle_{p,q}},$$

and compute the value of a specified loss function $R(\rho^{(i)}, \hat{a}_{p,q}^{(i)})$.

3. Choose the estimate of (ρ, a) which minimizes $R(\cdot)$.

5. PARAMETER ESTIMATES FROM MODELS BASED ON REGULARLY VARYING EXPRESSIONS FOR THE TAILS

5.1 Parametric Modeling For The Tail

A third approach to formulating tail estimates is to derive parametric models for the tails of the quantile function, distribution function, and density function from $fQ(1-u) = u^{\rho+1}L(u)$. Assuming these models hold beyond a given threshold, parameter estimates are derived from the largest order statistics.

The results of the following theorem motivate the probability models for the tails.

THEOREM 5.1.1. *Suppose that $fQ(1-u) = u^{\rho+1}L(u)$, where $\rho \in \mathbb{R}$ and $L(u)$ is slowly varying as $u \rightarrow 0^+$. Then*

(a) *the quantile function can be represented as*

(i) $Q(1-u) \sim Q(1) + u^{-\rho} \cdot L_0^*(u)$ as $u \rightarrow 0^+$ for $\rho < 0$;

(ii) $Q(1-u) \sim u^{-\rho} \cdot L_0^*(u)$ as $u \rightarrow 0^+$ for $\rho > 0$;

(b) *the distribution function can be represented as*

(i) $F[Q(1) - x] \sim 1 - x^{-1/\rho} \cdot L_1^*(x)$ as $x \rightarrow 0^+$ for $\rho < 0$ assuming $L(u[-\rho L(u)]^{1/\rho})/L(u) \rightarrow 1$ locally uniformly in $\rho < 0$ as $u \rightarrow 0^+$;

(ii) $F(x) \sim 1 - x^{-1/\rho} \cdot L_2^*(x)$ as $x \rightarrow \infty$ for $\rho > 0$ assuming $L(u)/L(u[\rho L(u)]^{1/\rho}) \rightarrow 1$ locally uniformly in $\rho > 0$ as $u \rightarrow 0^+$;

(c) *the density function can be represented as*

(i) $f[Q(1) - x] \sim (-1/\rho)x^{-(1/\rho)-1} \cdot L_1^*(x)$ as $x \rightarrow 0^+$ for $\rho < 0$ assuming $L(u[-\rho L(u)]^{1/\rho})/L(u) \rightarrow 1$ locally uniformly in $\rho < 0$ as $u \rightarrow 0^+$ and f is ultimately monotone;

(ii) $f(x) \sim (1/\rho)x^{-(1/\rho)-1} \cdot L_2^*(x)$ as $x \rightarrow \infty$ for $\rho > 0$ assuming $L(u)/L(u[\rho L(u)]^{1/\rho}) \rightarrow 1$ locally uniformly in $\rho > 0$ as $u \rightarrow 0^+$ and f is ultimately monotone;

where $L_0^*(u) = 1/[\rho L(u)]$, which is slowly varying as $u \rightarrow 0^+$, $L_1^*(x) = [-\rho L(x^{-1/\rho})]^{-1/\rho}$, which is slowly varying as $x \rightarrow 0^+$, and $L_2^*(x) =$

$[\rho L(x^{-1/\rho})]^{-1/\rho}$ which is slowly varying as $x \rightarrow \infty$.

Tables 3-5 contain, respectively, examples of the quantile function, distribution function, and density function for some common parametric probability models derived from the tail behavior model and expressed in the form given in Theorem 5.1.1.

Taking $L(u) = 1/\Delta$, $\Delta > 0$ for $u > t$ for some threshold percentile t suggests the tail parameterizations

$$Q(u) = \begin{cases} Q(1) - (1-u)^{-\rho} \cdot (\Delta/(-\rho)), & \rho < 0 \\ (1-u)^{-\rho} \cdot \Delta/\rho, & \rho > 0 \end{cases}$$

$$F(x) = \begin{cases} 1 - [Q(1) - x]^{-1/\rho} \cdot (-\rho/\Delta)^{-1/\rho}, & \rho < 0 \\ 1 - x^{-1/\rho} \cdot (\rho/\Delta)^{-1/\rho}, & \rho > 0 \end{cases}$$

$$f(x) = \begin{cases} (-1/\rho)[Q(1) - x]^{-(1/\rho)-1} \cdot (-\rho/\Delta)^{-1/\rho}, & \rho < 0 \\ (1/\rho)x^{-(1/\rho)-1} \cdot (\rho/\Delta)^{-1/\rho}, & \rho > 0 \end{cases}$$

Notice that this parameterization requires that the class of tail behavior is known. Each case will be treated separately in the following subsections.

Note that determining the class of tail behavior for the underlying distribution is equivalent to determining the domain of attraction of the extreme value distribution since

- (i) $\rho < 0 \Leftrightarrow$ Domain of Attraction is the Type III Extreme Value Distribution;
 - (ii) $\rho > 0 \Leftrightarrow$ Domain of Attraction is the Type II Extreme Value Distribution.
- Therefore, one method for determining the class of tail behavior is to evaluate the probability modeling assumptions and determine the domain of attraction.

A diagnostic derived from the sample which can be used to determine the class of tail behavior is the *Identification Quantile (IQ) Box Plot* defined by Parzen (1983). For an arbitrary random variable Y , define the *identification quantile standardized random variable* $ZQI = (Y - Q(.5))/\sigma$, where $Q(.5)$ is the median and $\sigma = 2[Q(.75) - Q(.25)]$ is the quartile deviation. Appendix A contains

TABLE 3

Quantile functions of common parametric probability models expressed in the representation given in Theorem 5.1.1(a) based on the tail behavior model $fQ(1-u) = u^{\rho+1}L(u)$.

Distribution	Quantile Function
Uniform	$Q(1-u) = .5 - u$ $= Q(1) + u \cdot (-1)$
Neg. Exponential	$Q(1-u) = [\ln 2 / (2 \ln 3)] + [\ln(1-u) / (2 \ln 3)]$ $\sim Q(1) + u \cdot (-1 / (2 \ln 3)) [1 + .5u] \quad \text{as } u \rightarrow 0^+$
Neg. Weibull(p)	$Q(1-u) = ([\ln 2]^{1/p} / \sigma) - ([-\ln(1-u)]^{1/p} / \sigma)$ $\sim Q(1) + u^{1/p} \cdot (-1/\sigma) [1 + .5(3 - p^{-1})u]$ $\text{as } u \rightarrow 0^+$
Cauchy	$Q(1-u) = \frac{1}{4} \tan \pi(\frac{1}{2} - u)$ $\sim u^{-1} \cdot (1/4\pi) [1 - (\pi^2/3)u^2] \quad \text{as } u \rightarrow 0^+$
Pareto(p)	$Q(1-u) = u^{-1/p} \cdot (1/\sigma) [1 - 2^{1/p} u^{1/p}]$
Fréchet(p)	$Q(1-u) = ([-\ln(1-u)]^{-1/p} / \sigma) - ([\ln 2]^{-1/p} / \sigma)$ $\sim u^{-1/p} \cdot (1/\sigma) [1 + .5(3 + p^{-1})u] \quad \text{as } u \rightarrow 0^+$

Note: See Appendix A for the definition of σ , a different scale constant for each different distribution.

TABLE 4

Distribution functions of common parametric probability models expressed in the representation given in Theorem 5.1.1(b) based on the tail behavior model $fQ(1-u) = u^{\rho+1}L(u)$.

Distribution	Distribution Function
Uniform	$F(Q(1) - x) = 1 - x \cdot I(0 < x < 1)$
Neg. Exponential	$F(Q(1) - x) = \begin{cases} 1, & x \leq 0 \\ e^{-(2 \ln 3)x}, & x > 0 \end{cases}$ $\sim 1 - x \cdot (2 \ln 3)[1 - (\ln 3)x] \quad \text{as } x \rightarrow 0^+$
Neg. Weibull(p)	$F(Q(1) - x) = \begin{cases} 1, & x \leq 0 \\ e^{-(\sigma x)^p}, & x > 0 \end{cases}$ $\sim 1 - x^p \cdot \sigma^p [1 + .5(1 - 3p)x^p]$ <p style="text-align: right;">as $x \rightarrow 0^+$</p>
Cauchy	$F(x) = .5 + (1/\pi) \tan^{-1} 4x$ $\sim 1 - x^{-1} \cdot (1/4\pi)[1 - (1/48)x^{-2}] \quad \text{as } x \rightarrow \infty$
Pareto(p)	$F(x) = \begin{cases} 0, & x \leq [1 - 2^{1/p}]/\sigma \\ 1 - [2^{1/p} + \sigma x]^{-p}, & x > [1 - 2^{1/p}]/\sigma \end{cases}$ $\sim 1 - x^{-p} \cdot \sigma^{-p} [1 - (p2^{1/p}/\sigma)x^{-1}]$ <p style="text-align: right;">as $x \rightarrow \infty$</p>
Fréchet(p)	$F(x) = \begin{cases} 0, & x \leq -(\ln 2)^{-1/p}/\sigma \\ \exp\{ -[(\ln 2)^{-1/p} + \sigma x]^{-p} \}, & x > -(\ln 2)^{-1/p}/\sigma \end{cases}$ $\sim 1 - x^{-p} \cdot \sigma^{-p} [1 - .5(1 + 3p)x^{-p}] \quad \text{as } x \rightarrow \infty$

Note: See Appendix A for the definition of σ , a different scale constant for each different distribution.

TABLE 5

Probability density functions of common parametric probability models expressed in the representation given in Theorem 5.1.1(c) based on the tail behavior model $fQ(1-u) = u^{\rho+1}L(u)$.

Distribution	Density Function
Uniform	$f(Q(1) - x) = x^0 \cdot I(0 < x < 1)$
Neg. Exponential	$f(Q(1) - x) = \begin{cases} (2 \ln 3)e^{-(2 \ln 3)x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$ $\sim x^0 \cdot (2 \ln 3)[1 - (2 \ln 3)x] \quad \text{as } x \rightarrow 0^+$
Neg. Weibull(p)	$f(Q(1) - x) = \begin{cases} p\sigma^p x^{p-1} e^{-(DQ x)^p}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$ $\sim px^{p-1} \cdot \sigma^p [1 + .5(1 - 3p)x^p] \quad \text{as } x \rightarrow 0^+$
Cauchy	$f(x) = (4/\pi) \cdot 1/(1 + 16x^2)$ $\sim x^{-2} \cdot (1/4\pi)[1 - (1/16)x^{-2}] \quad \text{as } x \rightarrow \infty$
Pareto(p)	$f(x) = \begin{cases} 0, & x \leq (1 - 2^{1/p})/\sigma \\ \sigma p [2^{1/p} + \sigma x]^{-p-1}, & x > (1 - 2^{1/p})/\sigma \end{cases}$ $\sim px^{-p-1} \cdot \sigma^{-p} [1 - (2^{1/p}(p+1)/\sigma)x^{-1}] \quad \text{as } x \rightarrow \infty$
Fréchet(p)	$f(x) = \begin{cases} 0, & x \leq -(\ln 2)^{-1/p}/\sigma \\ \sigma p [(\ln 2)^{-1/p} + \sigma x]^{-p-1} \cdot \exp\{ - [(\ln 2)^{-1/p} + \sigma x]^{-p} \}, & x > -(\ln 2)^{-1/p}/\sigma \end{cases}$ $\sim px^{-p-1} \cdot \sigma^{-p} [1 + .5(1 + 3p)x^{-p}] \quad \text{as } x \rightarrow \infty$

Note: See Appendix A for the definition of σ , a different scale constant for each different distribution.

many common parametric models expressed in the identification standardized version.

The standardized ZQI has identification quantile function $QI(u) = (Q(u) - Q(.5))/\sigma$. At $u = .5$, $QI(u)$ is equal to zero and has slope approximately one. This approximately tangent line is a basis for comparing the tail behavior of different distributions as $u \rightarrow 0^+$ or $u \rightarrow 1^-$. Figure 4 shows an overlay of the identification quantile functions for the Uniform distribution ($\rho < 0$), Normal distribution, and Cauchy distribution ($\rho > 0$), each of which is discussed in Appendix A. Notice from the figure that the three types of tail behavior are clearly differentiated as $u \rightarrow 1^-$.

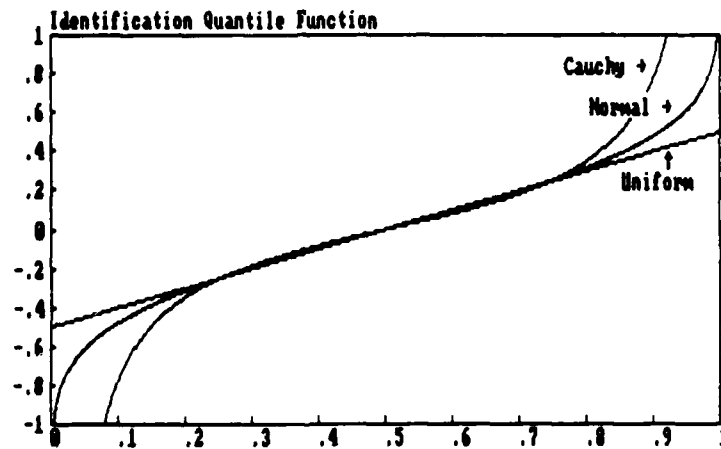


FIG. 4. Identification Quantile Plots for the Uniform ($\rho < 0$), Normal, and Cauchy ($\rho > 0$) Distributions clearly differentiating the types of tail behavior.

Therefore, a useful diagnostic for tail behavior is an estimate of the identification quantile function. Let

$$Q\Gamma(u) = \frac{Q^-(u) - Q^-(.5)}{2(Q^-(.75) - Q^-(.25))},$$

and display this function graphically in the IQ box plot. The IQ box plot is a graph of $Q\Gamma(u)$ for $0 < u < 1$ with informative overlays to help in using the plot as a diagnostic for classifying tail behavior.

The first of these overlays is used to indicate short tail behavior or equivalently $\rho < 0$. From the plots of the quantile functions of short tailed distributions

given in Appendix A, it is seen that, as a group, these quantile functions hardly deviate from the approximately tangent line at $u = .5$ with intercept zero and slope one (which is the identification quantile function of the Uniform distribution). Drawing this tangent line in the IQ box plot permits a visual diagnostic of short tail behavior.

Notice that the identification quantile of the Normal distribution drawn in Figure 4 is nearly equal to one at $u = 1$. It is useful to truncate the plots on $-1 \leq Q\Gamma(u) \leq 1$ to allow comparison to the Normal since it is an important special case in the family of parametric models and assumptions of normality are often made.

Long tail behavior or equivalently $\rho > 0$ is indicated when the sample identification quantile function exceeds the truncation line for values of u less than one.

5.2 Estimation Assuming The Class Of Tail Behavior Is Short Tailed

To obtain tail estimates when the tail behavior is known to be short tailed, choose, as a function of n , a threshold percentile t_n . Then let the threshold value be given by $T_n = Q^-(1 - t_n)$. Parameter estimates are obtained by assuming that the underlying distribution satisfies

$$(5.2.1) \quad Q(u) = Q(1) - (1 - u)^{-\rho} \cdot (\Delta / (-\rho)) \quad \text{for } t_n < u \leq 1$$

$$(5.2.2) \quad F(x) = 1 - [Q(1) - x]^{-1/\rho} \cdot (-\rho/\Delta)^{-1/\rho} \quad \text{for } Q^-(1 - t_n) < x \leq Q(1)$$

$$(5.2.3) \quad f(x) = (-1/\rho)[Q(1) - x]^{-(1/\rho)-1} \cdot (-\rho/\Delta)^{-1/\rho} \\ \text{for } Q^-(1 - t_n) < x \leq Q(1)$$

where $\rho < 0$, $Q(1) \in \mathbb{R}$, and $\Delta > 0$ are unknown parameter values.

Hall (1982) derives parameter estimates based on the largest order statistics. Stated in terms of the exceedence over threshold approach, the $[nt_n] + 1$ largest order statistics are given by

$$Q^-(1 - t_n) = T_n = X(n - [nt_n]; n) < X(n - [nt_n] + 1; n) < \cdots < X(n; n).$$

If the underlying distribution satisfies (5.2.1–5.2.3), the log-likelihood of the

$[nt_n] + 1$ largest order statistics is given by

$$\begin{aligned} \mathcal{L}_{ST}(\rho, Q(1), \Delta; X) &= \ln \frac{n!}{(n - [nt_n] - 1)!} \\ &\quad - ([nt_n] + 1) \left(\frac{1}{\rho} + 1 \right) \ln(-\rho) + ([nt_n] + 1) \frac{1}{\rho} \ln \Delta \\ &\quad - \left(\frac{1}{\rho} + 1 \right) \sum_{i=1}^{[nt_n]+1} \ln[Q(1) - X(n - i + 1; n)] \\ &\quad + (n - [nt_n] - 1) \ln \left(1 - \left[\frac{-\rho}{\Delta} [Q(1) - X(n - [nt_n]; n)] \right]^{-1/\rho} \right), \end{aligned}$$

where $X(n; n) < Q(1) < \infty$, $\rho < 0$, and $\Delta > 0$.

Solving $\partial \mathcal{L}_{ST}(\rho, Q(1), \Delta; X) / \partial \Delta = 0$ expresses the estimate $\hat{\Delta}$ given the parameter estimates $\hat{\rho}$ and $\hat{Q}(1)$:

$$\begin{aligned} \hat{\Delta} &= -\hat{\rho} [t_n + (1/n)]^{\hat{\rho}} [\hat{Q}(1) - X(n - [nt_n]; n)] \\ (5.2.4) \quad &\sim -\hat{\rho} t_n^{\hat{\rho}} [\hat{Q}(1) - X(n - [nt_n]; n)]. \end{aligned}$$

Hall (1982) uses this expression to reduce the three parameter log-likelihood \mathcal{L}_{ST} to a function of only ρ and $Q(1)$ by defining, for $X(n; n) < Q(1) < \infty$,

$$\begin{aligned} \mathcal{L}_{ST}^*(\rho, Q(1); X) &= \mathcal{L}_{ST}(\rho, Q(1), \hat{\Delta}; X) \\ &\sim \ln \frac{n!}{(n - [nt_n] - 1)!} t_n^{[nt_n]+1} (1 - t_n)^{n - [nt_n] - 1} \\ &\quad - ([nt_n] + 1) \ln(-\rho) + ([nt_n] + 1) \frac{1}{\rho} \ln[Q(1) - X(n - [nt_n]; n)] \\ &\quad - \left(\frac{1}{\rho} + 1 \right) \sum_{i=1}^{[nt_n]+1} \ln[Q(1) - X(n - i + 1; n)], \end{aligned}$$

which is the same function (up to terms not involving ρ and $Q(1)$) maximized by Smith and Weissman (1985) to obtain parameter estimates.

If $\rho < -1$, the function $\mathcal{L}_{ST}^*(\cdot)$ has no maximum since for any $\rho < -1$, $\lim_{Q(1) \rightarrow X(n; n)+} \mathcal{L}_{ST}^*(\rho, Q(1); X) = \infty$. Therefore, Hall's estimates based on the largest order statistics from a short tailed probability model are denoted

by $(\hat{\rho}, \hat{Q}(1), \hat{\Delta})$, where $(\hat{\rho}, \hat{Q}(1))$ is the maximum of $\mathcal{L}_{ST}^*(\rho, Q(1); X)$ in the constrained parameter space $\{-1 < \rho < 0, X(n; n) < Q(1) < \infty\}$, and $\hat{\Delta}$ is given by (5.2.4).

To derive the large sample properties of Hall's estimators for the short tailed case based on the largest order statistics, first express Hall's estimators for ρ and $Q(1)$ as solutions to a set of estimating equations, take the Taylor's series expansion, and then compute the asymptotic distribution of each term. The distribution of $\hat{\Delta}$ then follows since it is a function of $\hat{\rho}$ and $\hat{Q}(1)$. This approach yields the following result.

THEOREM 5.2.1. *Suppose that $fQ(1-u) = u^{\rho+1}L(u)$, where $-\frac{1}{2} < \rho < 0$ and*

$$\left| \frac{L(t)}{L(tu)} - 1 \right| \leq A(u)R(t)$$

for some positive measurable functions $A(u)$ and $R(t)$ where $\lim_{t \rightarrow 0+} R(t) = 0$. Let $\{T_n\}$ be a sequence of thresholds defined on $(Q(0), Q(1))$ such that $nt_n^ \rightarrow \infty$ and $t_n^* \rightarrow 0$ as $n \rightarrow \infty$, where $t_n^* = 1 - F(T_n)$. Further, suppose $-\rho[Q(1) - T_n]hQ(1 - t_n^*) = 1 + O(R_1(t_n^*))$ for some positive measurable function $R_1(t)$ with $\lim_{t \rightarrow 0+} R_1(t) = 0$. Let $(\hat{\rho}_n, \hat{Q}_n(1), \hat{\Delta}_n)$ denote Hall's estimates based on the largest order statistics from a short tailed probability model. Then conditional on T_n ,*

$$\begin{bmatrix} \hat{\rho}_n \\ \hat{Q}_n(1) \\ \hat{\Delta}_n \end{bmatrix} \text{ is AN } \left(\begin{bmatrix} \rho + O(R^*(t_n^*)) \\ Q(1) + O(R^*(t_n^*)) \\ -\rho(t_n^*)^\rho[Q(1) - T_n] + O(R^*(t_n^*)) \end{bmatrix}, ([nt_n^*])^{-1}V_{ST} \right)$$

as $nt_n^ \rightarrow \infty$, where $R^*(t) = \max\{R(t), R_1(t)\}$ and*

$$V_{ST} = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix},$$

with

$$v_{11} = (\rho + 1)^2 + O(R^*(t_n^*)),$$

$$v_{22} = \left(\frac{1}{\rho} + 1\right)^2 (2\rho + 1)[Q(1) - T_n]^2 + O(R^*(t_n^*)),$$

$$v_{33} = \rho^2(\rho + 1)[Q(1) - T_n]^2(t_n^*)^\rho[(\rho + 1)(\ln t_n^*)^2 - 2 \ln t_n^* + 2] + O(R^*(t_n^*)),$$

$$v_{12} = v_{21}$$

$$= -\left(\frac{1}{\rho} + 1\right) (2\rho + 1)[Q(1) - T_n] + O(R^*(t_n^*)),$$

$$v_{13} = v_{31}$$

$$= \rho(\rho + 1)[Q(1) - T_n](t_n^*)^\rho[1 - (\rho + 1) \ln t_n^*] + O(R^*(t_n^*)),$$

$$v_{23} = v_{32}$$

$$= (\rho + 1)(2\rho + 1)[Q(1) - T_n]^2(t_n^*)^\rho(\ln t_n^* - 1) + O(R^*(t_n^*)).$$

PROOF. Attention will first be paid to the pair $(\hat{\rho}, \hat{Q}(1))$ since $\hat{\Delta}$ is a function of these two parameters. Write

$$-\frac{\partial \mathcal{L}_{ST}^*(\rho, Q(1); X)}{\partial \rho}$$

and

$$-\frac{\partial \mathcal{L}_{ST}^*(\rho, Q(1); X)}{\partial Q(1)}$$

as sums of the independent exceedences $X_i - T_n$ given $X_i > T_n$. That is,

$$\begin{aligned} -\frac{\partial \mathcal{L}_{ST}^*(\rho, Q(1); X)}{\partial \rho} &= \frac{[nt_n^*] + 1}{\rho} - \frac{1}{\rho^2} \sum_{i=1}^{[nt_n^*]} \ln \left[1 - \frac{X(n-i+1; n) - T_n}{Q(1) - T_n} \right] \\ &= \sum_{i=1}^{[nt_n^*]} \left\{ \frac{1}{\rho} \left(1 + \frac{1}{[nt_n^*]} \right) - \frac{1}{\rho^2} \ln \left(1 - \frac{X_i - T_n}{Q(1) - T_n} \right) \right\} \\ &= \sum_{i=1}^{[nt_n^*]} \psi_1(\rho, Q(1); X_i - T_n) \\ -\frac{\partial \mathcal{L}_{ST}^*(\rho, Q(1); X)}{\partial Q(1)} &= \frac{[nt_n^*]}{\rho[Q(1) - T_n]} \\ &\quad - \left(\frac{1}{\rho} + 1 \right) \frac{1}{Q(1) - T_n} \sum_{i=1}^{[nt_n^*]} \left[1 - \frac{X(n-i+1; n) - T_n}{Q(1) - T_n} \right]^{-1} \\ &= \sum_{i=1}^{[nt_n^*]} \left\{ \frac{1}{\rho[Q(1) - T_n]} \right\} \end{aligned}$$

$$\begin{aligned}
& - \left(\frac{1}{\rho} + 1 \right) \frac{1}{Q(1) - T_n} \left[1 - \frac{X_i - T_n}{Q(1) - T_n} \right]^{-1} \Big\} \\
& = \sum_{i=1}^{[nt_n^*]} \psi_2(\rho, Q(1); X_i - T_n)
\end{aligned}$$

Therefore, Hall's estimates $(\hat{\rho}_n, \hat{Q}_n(1))$ can be expressed as the solution to the estimating equation

$$0 = ([nt_n^*])^{-1/2} \sum_{i=1}^{[nt_n^*]} \begin{bmatrix} \psi_1(\hat{\rho}_n, \hat{Q}_n(1); X_i - T_n) \\ \psi_2(\hat{\rho}_n, \hat{Q}_n(1); X_i - T_n) \end{bmatrix}.$$

Take the first order Taylor's series expansion about the true parameter values $(\rho, Q(1))$ of the right hand side to obtain, for some point $(\rho^*, Q^*(1))$ on the line segment between $(\hat{\rho}_n, \hat{Q}_n(1))$ and $(\rho, Q(1))$,

$$\begin{aligned}
0 &= ([nt_n^*])^{-1/2} \sum_{i=1}^{[nt_n^*]} \begin{bmatrix} \psi_1(\rho, Q(1); X_i - T_n) \\ \psi_2(\rho, Q(1); X_i - T_n) \end{bmatrix} \\
&+ ([nt_n^*])^{-1} \sum_{i=1}^{[nt_n^*]} \begin{bmatrix} -\frac{\partial \psi_1(\rho^*, Q^*(1); X_i - T_n)}{\partial \rho} & -\frac{\partial \psi_1(\rho^*, Q^*(1); X_i - T_n)}{\partial Q(1)} \\ -\frac{\partial \psi_2(\rho^*, Q^*(1); X_i - T_n)}{\partial \rho} & -\frac{\partial \psi_2(\rho^*, Q^*(1); X_i - T_n)}{\partial Q(1)} \end{bmatrix} \\
&\cdot ([nt_n^*])^{1/2} \begin{bmatrix} \hat{\rho}_n - \rho \\ \hat{Q}_n(1) - Q(1) \end{bmatrix}.
\end{aligned}$$

From the Central Limit Theorem for an iid sequence,

$$([nt_n^*])^{-1/2} \sum_{i=1}^{[nt_n^*]} \begin{bmatrix} \psi_1(\rho, Q(1); X_i - T_n) \\ \psi_2(\rho, Q(1); X_i - T_n) \end{bmatrix} \quad \text{is} \quad \text{AN}(\mu_n, ([nt_n^*])^{-1} \Sigma_n)$$

as $nt_n^* \rightarrow \infty$, where

$$\begin{aligned} \mu_n &= E \begin{bmatrix} X - T_n \\ X > T_n \end{bmatrix} \begin{bmatrix} \psi_1(\rho, Q(1); X_i - T_n) \\ \psi_2(\rho, Q(1); X_i - T_n) \end{bmatrix} \\ &= \begin{bmatrix} (\rho[nt_n^*])^{-1} + O(R^*(t_n^*)) \\ O(R^*(t_n^*)) \end{bmatrix} \sim \begin{bmatrix} O(R^*(t_n^*)) \\ O(R^*(t_n^*)) \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} \Sigma_n &= \text{Cov} \begin{bmatrix} X - T_n \\ X > T_n \end{bmatrix} \begin{bmatrix} \psi_1(\rho, Q(1); X_i - T_n) \\ \psi_2(\rho, Q(1); X_i - T_n) \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}, \end{aligned}$$

with

$$\begin{aligned} \sigma_{11} &= \frac{1}{\rho^2} \left(1 - \frac{1}{[nt_n^*]} \right) + O(R^*(t_n^*)) \\ &\sim \frac{1}{\rho^2} + O(R^*(t_n^*)), \\ \sigma_{22} &= \left(\frac{1}{2\rho + 1} + \frac{1}{\rho[nt_n^*]} \right) [Q(1) - T_n]^{-2} + O(R^*(t_n^*)) \\ &\sim \frac{1}{2\rho + 1} [Q(1) - T_n]^{-2} + O(R^*(t_n^*)), \\ \sigma_{12} &= \sigma_{21} \\ &= \frac{1}{\rho(\rho + 1)} [Q(1) - T_n]^{-1} + O(R^*(t_n^*)). \end{aligned}$$

This result follows from the expressions for the moments given in Theorem 2.4.1.

For n sufficiently large to make the $O(R^*(t_n^*))$ terms negligible, notice that Σ_n is positive definite if and only if $-\frac{1}{2} < \rho < 0$.

If Σ_n is positive definite, then from the Weak Law of Large Numbers and the fact that $(\rho^*, Q^*(1)) \xrightarrow{p} (\rho, Q(1))$ as $nt_n^* \rightarrow \infty$, it follows that

$$\Sigma_n^{-1} \cdot ([nt_n^*])^{-1} \sum_{i=1}^{[nt_n^*]} \begin{bmatrix} -\frac{\partial \psi_1(\rho^*, Q^*(1); X_i - T_n)}{\partial \rho} & -\frac{\partial \psi_1(\rho^*, Q^*(1); X_i - T_n)}{\partial Q(1)} \\ -\frac{\partial \psi_2(\rho^*, Q^*(1); X_i - T_n)}{\partial \rho} & -\frac{\partial \psi_2(\rho^*, Q^*(1); X_i - T_n)}{\partial Q(1)} \end{bmatrix}$$

$$\xrightarrow{p} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

as $nt_n^* \rightarrow \infty$. This result follows from the expressions of the derivatives for the estimating equations given in Appendix F and the moments given in Theorem 2.4.1.

The asymptotic distribution of $(\hat{\rho}_n, \hat{Q}_n(1))$ then follows from Slutsky's Theorem. Since $\hat{\Delta}_n$ is a function of $(\hat{\rho}_n, \hat{Q}_n(1))$, the result of the theorem follows.

□

It is not immediately obvious that the tail estimates motivated from the parameterization in this section fit into that given in (2.3.5–2.3.7). However, notice that for $1 - t_n^* < u \leq 1$,

$$\begin{aligned} Q^*(u) &= Q^*(1) - (1 - u)^{-\hat{\rho}} \cdot (\hat{\Delta}/(-\hat{\rho})) \\ &= Q^*(1) - (1 - u)^{-\hat{\rho}} \left(\frac{-\hat{\rho}(t_n^*)^{\hat{\rho}}[Q^*(1) - T_n]}{\hat{\rho}} \right) \\ &= Q^*(1) - \left(\frac{1 - u}{t_n^*} \right)^{-\hat{\rho}} [Q^*(1) - T_n] \\ &= T_n - [Q^*(1) - T_n] \left[\left(\frac{1 - u}{t_n^*} \right)^{-\hat{\rho}} - 1 \right] \\ &= T_n + (-\hat{\rho}[Q^*(1) - T_n]) \cdot \left[-g \left(\frac{1 - u}{t_n^*}; -\hat{\rho} \right) \right], \end{aligned}$$

and for $T_n < x \leq Q(1)$,

$$F^*(x) = 1 - [Q^*(1) - x]^{-1/\hat{\rho}} \cdot (-\hat{\rho}/\hat{\Delta})^{-1/\hat{\rho}}$$

$$\begin{aligned}
&= 1 - [Q^*(1) - x]^{-1/\hat{\rho}} \left[\frac{-\hat{\rho}}{-\hat{\rho}(t_n^*)^{\hat{\rho}}[Q^*(1) - T_n]} \right]^{-1/\hat{\rho}} \\
&= 1 - t_n^* \left[\frac{Q^*(1) - x}{Q^*(1) - T_n} \right]^{-1/\hat{\rho}} \\
&= 1 - t_n^* \left[1 - \frac{x - T_n}{Q^*(1) - T_n} \right]^{-1/\hat{\rho}} \\
&= 1 - t_n^* \cdot g^{-1} \left(\frac{1}{\hat{\rho}[Q^*(1) - T_n]}(x - T_n); -\hat{\rho} \right),
\end{aligned}$$

and

$$\begin{aligned}
f^*(x) &= \left(-\frac{1}{\hat{\rho}} \right) [Q^*(1) - x]^{-(1/\hat{\rho})-1} \cdot (-\hat{\rho}/\hat{\Delta})^{-1/\hat{\rho}} \\
&= \left(-\frac{1}{\hat{\rho}} \right) [Q^*(1) - x]^{-(1/\hat{\rho})-1} \left[\frac{-\hat{\rho}}{-\hat{\rho}(t_n^*)^{\hat{\rho}}[Q^*(1) - T_n]} \right]^{-1/\hat{\rho}} \\
&= t_n^* \left(\frac{1}{-\hat{\rho}[Q^*(1) - T_n]} \right) \left[1 - \frac{x - T_n}{Q^*(1) - T_n} \right]^{-(1/\hat{\rho})-1} \\
&= t_n^* \left(\frac{1}{-\hat{\rho}[Q^*(1) - T_n]} \right) \cdot (g^{-1})' \left(\frac{1}{\hat{\rho}[Q^*(1) - T_n]}(x - T_n); -\hat{\rho} \right).
\end{aligned}$$

Therefore, the tail estimates based on the parameter estimates from this parameterization of the tail follow under the unified approach proposed in Subsection 2.3 where $\hat{a} = -\hat{\rho}[Q^*(1) - T_n]$. For comparison with other parameter estimates, the following corollary is easily shown.

COROLLARY 5.2.1. *Suppose the assumptions of Theorem 5.2.1 hold. Then conditional on T_n ,*

$$\begin{aligned}
\begin{bmatrix} \hat{\rho}_n \\ \hat{a}_n \end{bmatrix} &\text{ is AN } \left(\begin{bmatrix} \rho + O(R^*(t_n^*)) \\ -\rho[Q(1) - T_n] + O(R^*(t_n^*)) \end{bmatrix} = \begin{bmatrix} \rho + O(R^*(t_n^*)) \\ a + O(R^*(t_n^*)) \end{bmatrix}, \right. \\
&\quad \left. ([nt_n^*])^{-1} V_{ST}^* \right)
\end{aligned}$$

as $nt_n^* \rightarrow \infty$, where

$$V_{ST}^* = \begin{bmatrix} v_{11}^* & v_{12}^* \\ v_{21}^* & v_{22}^* \end{bmatrix},$$

with

$$\begin{aligned}
 v_{11}^* &= (\rho + 1)^2 + O(R^*(t_n^*)), \\
 v_{22}^* &= 2\rho^2(\rho + 1)[Q(1) - T_n]^2 + O(R^*(t_n^*)) \\
 &= 2a^2(\rho + 1) + O(R^*(t_n^*)), \\
 v_{12}^* &= v_{21}^* \\
 &= \rho(\rho + 1)[Q(1) - T_n] + O(R^*(t_n^*)) \\
 &= -a(\rho + 1) + O(R^*(t_n^*)).
 \end{aligned}$$

5.3 Estimation Assuming The Class Of Tail Behavior Is Long Tailed

To obtain tail estimates when the tail behavior is known to be long tailed, choose, as a function of n , a threshold percentile t_n . Then let the threshold value be given by $T_n = Q^-(1 - t_n)$. Parameter estimates are obtained by assuming that the underlying distribution satisfies

$$(5.3.1) \quad Q(u) = (1 - u)^{-\rho} \cdot \Delta / \rho, \quad 1 - t_n < u < 1,$$

$$(5.3.2) \quad F(x) = 1 - x^{-1/\rho} \cdot (\rho / \Delta)^{-1/\rho}, \quad x > Q^-(1 - t_n),$$

$$(5.3.3) \quad f(x) = (1/\rho)x^{-(1/\rho)-1} \cdot (\rho / \Delta)^{-1/\rho}, \quad x > Q^-(1 - t_n),$$

where $\rho > 0$ and $\Delta > 0$ are unknown parameter values.

The most popular estimate of the tail exponent ρ was proposed by Hill (1975), who derived parameter estimates based on the largest order statistics. Stated in terms of the exceedence over threshold approach the $[nt_n] + 1$ largest order statistics are given by

$$Q^-(1 - t_n) = T_n = X(n - [nt_n]; n) < X(n - [nt_n] + 1; n) < \cdots < X(n; n).$$

If the underlying distribution satisfies (5.3.1–5.3.3), the log-likelihood of the $[nt_n] + 1$ largest order statistics is given by

$$\mathcal{L}_{LT}(\rho, \Delta; X) = \ln \frac{n!}{(n - [nt_n] - 1)!}$$

$$\begin{aligned}
& - ([nt_n] + 1) \left(\frac{1}{\rho} + 1 \right) \ln \rho + ([nt_n] + 1) \frac{1}{\rho} \ln \Delta \\
& - \left(\frac{1}{\rho} + 1 \right) \sum_{i=1}^{[nt_n]+1} \ln X(n - i + 1; n) \\
& + (n - [nt_n] - 1) \ln \left(1 - \left[\frac{\rho}{\Delta} X(n - [nt_n]; n) \right]^{-1/\rho} \right)
\end{aligned}$$

where $\rho > 0$, and $\Delta > 0$.

Solving $\partial \mathcal{L}_{LT}(\rho, \Delta; X) / \partial \Delta = 0$ expresses the estimate $\hat{\Delta}$ given the parameter estimate $\hat{\rho}$:

$$\begin{aligned}
\hat{\Delta} &= \hat{\rho} [t_n + (1/n)]^{\hat{\rho}} X(n - [nt_n]; n) \\
(5.3.4) \quad &\sim \hat{\rho} t_n^{\hat{\rho}} X(n - [nt_n]; n).
\end{aligned}$$

As in the maximization of the log-likelihood for the order statistics from a short tailed distribution, the two parameter log-likelihood \mathcal{L}_{LT} is reduced to a function of only ρ by defining

$$\begin{aligned}
\mathcal{L}_{LT}^*(\rho; X) &= \mathcal{L}_{LT}(\rho, \hat{\Delta}; X) \\
&\sim \ln \frac{n!}{(n - [nt_n] - 1)!} t_n^{[nt_n]+1} (1 - t_n)^{n - [nt_n] - 1} \\
&\quad - ([nt_n] + 1) \ln \rho + ([nt_n] + 1) \frac{1}{\rho} \ln T_n \\
&\quad - \left(\frac{1}{\rho} + 1 \right) \sum_{i=1}^{[nt_n]+1} \ln X(n - i + 1; n).
\end{aligned}$$

Therefore, Hill's estimate for the tail exponent ρ is found by solving $\partial \mathcal{L}_{LT}^*(\rho; X) / \partial \rho = 0$ to obtain

$$\begin{aligned}
\hat{\rho} &= \frac{1}{[nt_n] + 1} \left[\sum_{i=1}^{[nt_n]+1} \ln X(n - i + 1; n) - \ln X(n - [nt_n]; n) \right] \\
&= \frac{1}{[nt_n] + 1} \left[\sum_{i=1}^{[nt_n]} \ln X(n - i + 1; n) - \ln X(n - [nt_n]; n) \right] \\
(5.3.5) \quad &\sim \frac{1}{[nt_n]} \sum_{i=1}^{[nt_n]} \ln \left[\frac{X(n - i + 1; n)}{X(n - [nt_n]; n)} \right].
\end{aligned}$$

The asymptotic distribution of Hill's estimate has been considered by many authors. For example, Mason (1982) gives necessary and sufficient conditions for Hill's estimate to converge almost surely or in probability to a constant; Davis and Resnick (1984) motivate Hill's estimator from extreme value theory and derive its asymptotic normality.

The following theorem on the asymptotic normality of $\hat{\rho}$ is similar to one given by Goldie and Smith (1987), who also place rates of convergence on the slowly varying function and derive the asymptotic normality of Hill's estimate. The key difference in the theorem given here and Goldie and Smith's result is that the bias due to the threshold value is expressed in terms of rates of convergence. Also different from Goldie and Smith is the proof, which uses (2.2.1), the representation for the quantile function of the exceedences.

THEOREM 5.3.1. *Suppose that $fQ(1-u) = u^{\rho+1}L(u)$, where $\rho > 0$ and*

$$\left| \frac{L(t)}{L(tu)} - 1 \right| \leq A(u)R(t)$$

for some positive measurable functions $A(u)$ and $R(t)$ where $\lim_{t \rightarrow 0+} R(t) = 0$. Let $\{T_n\}$ be a sequence of thresholds defined on $(Q(0), Q(1))$ such that $nt_n^ \rightarrow \infty$ and $t_n^* \rightarrow 0$ as $n \rightarrow \infty$, where $t_n^* = 1 - F(T_n)$. Further, suppose $\rho T_n hQ(1-t_n^*) = 1 + O(R_1(t_n^*))$ for some positive measurable function $R_1(t)$ with $\lim_{t \rightarrow 0+} R_1(t) = 0$. Let $(\hat{\rho}_n, \hat{\Delta}_n)$ denote Hill's estimates based on the largest order statistics from a long tailed probability model given in (5.3.5) and (5.3.4). Then conditional on T_n ,*

$$\begin{bmatrix} \hat{\rho}_n \\ \hat{\Delta}_n \end{bmatrix} \text{ is AN } \left(\begin{bmatrix} \rho + O(R^*(t_n^*)) \\ \rho(t_n^*)^\rho T_n + O(R^*(t_n^*)) \end{bmatrix}, ([nt_n^*])^{-1} V_{LT} \right)$$

as $nt_n^ \rightarrow \infty$, where $R^*(t) = \max\{R(t), R_1(t)\}$ and*

$$V_{LT} = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix},$$

with

$$v_{11} = \rho^2 + O(R^*(t_n^*))$$

$$v_{22} = \rho^2 (t_n^*)^{2\rho} (\rho \ln t_n^* + 1)^2 T_n^2 + O(R^*(t_n^*))$$

$$v_{12} = v_{21}$$

$$= \rho^2 (t_n^*)^\rho (\rho \ln t_n^* + 1) T_n + O(R^*(t_n^*))$$

PROOF. Attention will first be paid to the distribution of $\hat{\rho}$ since $\hat{\Delta}$ is a function of ρ . First, write $\hat{\rho}$ as the sum of the independent exceedences $X_i - T_n$ given $X_i > T_n$,

$$\begin{aligned} \hat{\rho} &= ([nt_n^*])^{-1} \sum_{i=1}^{[nt_n^*]} \ln \frac{X(n-i+1; n)}{X(n-[nt_n^*]; n)} \\ &= ([nt_n^*])^{-1} \sum_{i=1}^{[nt_n^*]} \ln \left(1 + \frac{X_i - T_n}{T_n} \right). \end{aligned}$$

From the Central Limit Theorem for an iid sequence,

$$([nt_n^*])^{-1/2} \sum_{i=1}^{[nt_n^*]} \ln \left(1 + \frac{X_i - T_n}{T_n} \right) \quad \text{is} \quad \text{AN}(\mu_n, ([nt_n^*])^{-1} \sigma_n^2),$$

as $nt_n^* \rightarrow \infty$, where

$$\begin{aligned} \mu_n &= E_{X-T_n} \mid X > T_n \ln \left(1 + \frac{X_i - T_n}{T_n} \right) \\ &= \rho + O(R^*(t_n^*)) \end{aligned}$$

and

$$\begin{aligned} \sigma_n^2 &= \text{Var}_{X-T_n} \mid X > T_n \ln \left(1 + \frac{X_i - T_n}{T_n} \right) \\ &= \rho^2 + O(R^*(t_n^*)). \end{aligned}$$

This result follows from the expressions for the moments given in Theorem 2.4.1.

Since $\hat{\Delta}_n$ is a function of $\hat{\rho}_n$, the result of the theorem follows. \square

It is not immediately obvious that the tail estimates motivated from the parameterization in this section fit into that given in (2.3.5–2.3.7). However,

notice that for $1 - t_n^* < u < 1$,

$$\begin{aligned}
 Q^*(u) &= (1 - u)^{-\hat{\rho}} \cdot (\hat{\Delta}/\hat{\rho}) \\
 &= (1 - u)^{-\hat{\rho}} \left(\frac{\hat{\rho}(t_n^*)^{\hat{\rho}} T_n}{\hat{\rho}} \right) \\
 &= \left(\frac{1 - u}{t_n^*} \right)^{-\hat{\rho}} T_n \\
 &= T_n + T_n \left[\left(\frac{1 - u}{t_n^*} \right)^{-\hat{\rho}} - 1 \right] \\
 &= T_n + (\hat{\rho} T_n) \cdot \left[-g \left(\frac{1 - u}{t_n^*}; -\hat{\rho} \right) \right],
 \end{aligned}$$

and for $x > T_n$,

$$\begin{aligned}
 F^*(x) &= 1 - x^{-1/\hat{\rho}} \cdot (\hat{\rho}/\hat{\Delta})^{-1/\hat{\rho}} \\
 &= 1 - x^{-1/\hat{\rho}} \left[\frac{\hat{\rho}}{\hat{\rho}(t_n^*)^{\hat{\rho}} T_n} \right]^{-1/\hat{\rho}} \\
 &= 1 - t_n^* \left[\frac{x}{T_n} \right]^{-1/\hat{\rho}} \\
 &= 1 - t_n^* \left[1 + \frac{x - T_n}{T_n} \right]^{-1/\hat{\rho}} \\
 &= 1 - t_n^* \cdot g^{-1} \left(-\frac{1}{\hat{\rho} T_n} (x - T_n); -\hat{\rho} \right),
 \end{aligned}$$

and

$$\begin{aligned}
 f^*(x) &= \left(\frac{1}{\hat{\rho}} \right) x^{-(1/\hat{\rho})-1} \cdot (\hat{\rho}/\hat{\Delta})^{-1/\hat{\rho}} \\
 &= \left(\frac{1}{\hat{\rho}} \right) x^{-(1/\hat{\rho})-1} \left[\frac{\hat{\rho}}{\hat{\rho}(t_n^*)^{\hat{\rho}} T_n} \right]^{-1/\hat{\rho}} \\
 &= t_n^* \left(\frac{1}{\hat{\rho} T_n} \right) \left[1 + \frac{x - T_n}{T_n} \right]^{-(1/\hat{\rho})-1} \\
 &= t_n^* \left(\frac{1}{\hat{\rho} T_n} \right) \cdot (g^{-1})' \left(-\frac{1}{\hat{\rho} T_n} (x - T_n); -\hat{\rho} \right).
 \end{aligned}$$

Therefore, the tail estimates based on the parameter estimates from this parameterization of the tail follow under the unified approach proposed in Subsection 2.3 where $\hat{a} = \hat{\rho} T_n$. For comparison with other parameter estimates, the following corollary is easily shown.

COROLLARY 5.3.1. *Suppose the assumptions of Theorem 5.3.1 hold. Then conditional on T_n ,*

$$\begin{bmatrix} \hat{\rho}_n \\ \hat{a}_n \end{bmatrix} \text{ is AN } \left(\begin{bmatrix} \rho + O(R^*(t_n^*)) \\ \rho T_n + O(R^*(t_n^*)) \end{bmatrix} = \begin{bmatrix} \rho + O(R^*(t_n^*)) \\ a + O(R^*(t_n^*)) \end{bmatrix}, ([nt_n^*])^{-1} V_{LT}^* \right)$$

as $nt_n^* \rightarrow \infty$, where

$$\begin{aligned} V_{LT}^* &= \begin{bmatrix} \rho^2 + O(R^*(t_n^*)) & \rho^2 T_n + O(R^*(t_n^*)) \\ \rho^2 T_n + O(R^*(t_n^*)) & \rho^2 T_n^2 + O(R^*(t_n^*)) \end{bmatrix} \\ &= \begin{bmatrix} \rho^2 + O(R^*(t_n^*)) & a\rho + O(R^*(t_n^*)) \\ a\rho + O(R^*(t_n^*)) & a^2 + O(R^*(t_n^*)) \end{bmatrix}. \end{aligned}$$

6. DISCUSSION OF PARAMETER ESTIMATE PROPERTIES

6.1 Comparison Of Parameter Estimates

All comparisons are made assuming the sequence of threshold values $\{T_n\}$ satisfies $nt_n^* \rightarrow \infty$ as $n \rightarrow \infty$, where $t_n^* = 1 - F(T_n)$. This assumption is necessary for the asymptotic normality of the proposed parameter estimates. The tail exponent estimators are the focus of this discussion since the estimates of α show no important differences.

All of the parameter estimates proposed in Sections 3–5 are biased in general. This is due to the difference between the true value for the tail and the model for the tail used to obtain parameter estimates. Under the conditions of Theorems 3.2.1, 4.2.1, 5.2.1, 5.3.1,

$$\text{Bias}(\hat{\rho}_{\text{GPD}}) = O(R^*(t_n^*)), \quad -\frac{1}{2} < \rho < \infty, \rho \neq 0,$$

$$\text{Bias}(\hat{\rho}_{\text{GEV}}) = -\frac{\rho - 4}{4\rho(\rho - 2)^2} + O(R^*(t_n^*)), \quad -.356967 \approx \rho_0 < \rho < 1, \rho \neq 0,$$

$$\text{Bias}(\hat{\rho}_{\text{Hall}}) = O(R^*(t_n^*)), \quad -\frac{1}{2} < \rho < 0,$$

$$\text{Bias}(\hat{\rho}_{\text{Hill}}) = O(R^*(t_n^*)), \quad \rho > 0,$$

as $nt_n^* \rightarrow \infty$.

Only the GEV estimate fails to be asymptotically unbiased under the additional condition $t_n^* \rightarrow 0$ as $n \rightarrow \infty$. Therefore, the GEV estimate would be inappropriate for very large samples. A comparison of the magnitudes of the bias for small samples is a topic for further investigation.

No global statements can be made regarding the GPD estimates, Hall's estimates, and Hill's estimates since they are not defined on a common range of support. Assuming the underlying probability model is short tailed, the GPD

estimate and Hall's estimate are asymptotically equivalent since

$$\begin{aligned} \text{ARE}(\hat{\rho}_{\text{GPD}}, \hat{\rho}_{\text{Hall}}) &= \lim_{n \rightarrow \infty} \frac{\text{Var}(\hat{\rho}_{\text{GPD}})}{\text{Var}(\hat{\rho}_{\text{Hall}})} \\ &= \lim_{n \rightarrow \infty} \frac{([nt_n^*])^{-1} \{(\rho + 1)^2 + O(R^*(t_n^*))\}}{([nt_n^*])^{-1} \{(\rho + 1)^2 + O(R^*(t_n^*))\}} \\ &= 1, \quad -\frac{1}{2} < \rho < 0. \end{aligned}$$

Assuming the underlying probability model is long tailed, Hill's estimate is preferred since

$$\begin{aligned} \text{ARE}(\hat{\rho}_{\text{GPD}}, \hat{\rho}_{\text{Hill}}) &= \lim_{n \rightarrow \infty} \frac{\text{Var}(\hat{\rho}_{\text{GPD}})}{\text{Var}(\hat{\rho}_{\text{Hill}})} \\ &= \lim_{n \rightarrow \infty} \frac{([nt_n^*])^{-1} \{(\rho + 1)^2 + O(R^*(t_n^*))\}}{([nt_n^*])^{-1} \{\rho^2 + O(R^*(t_n^*))\}} \\ &= \left(1 + \frac{1}{\rho}\right)^2 \\ &> 1, \quad \rho > 0. \end{aligned}$$

This gives an example of the improved estimation possible through incorporating additional assumptions into the probability model. The IQ box plot discussed in Subsection 5.1 is one data analytic tool for validating assumptions on tail behavior.

6.2 Interpretation Of Parameter Estimates

The majority of the interest in tail behavior is in the tail exponent ρ because it describes important properties of the underlying distribution. For example, many statistical estimators make the assumption that the underlying distribution has finite variance, which corresponds to assuming $\rho < \frac{1}{2}$. Also of interest is the existence of a finite upper endpoint for the underlying distribution, which corresponds to assuming the short tailed class where $\rho < 0$. Davison (1984) proposes modeling the tail exponent as $\rho = \beta'x_j$ for a vector of design constants or covariates in order to allow comparisons of the tail behavior for different populations or pool information from related populations.

The focus of attention on the parameter estimates for interpretation is natural, but for the most part unjustified. The confusion over parameter estimate interpretation is twofold. First, the parameters defining tail behavior are non-identifiable and cannot be estimated without bias. Second, the true parameter value may not provide the best fit to the observed tail over the range of interest.

6.2.1 Parameters Are Nonidentifiable. One important complication to the interpretation of the parameter estimates is due to the bias in the estimates. This bias is not due to any failure of the methodology. Instead, it is due to the failure of a parametric model to completely specify the underlying distribution.

Using the estimate of ρ from the GPD modeling or either Hall's estimate or Hill's estimate depending on the appropriate classification of the underlying distribution, then asymptotically, $E(\hat{\rho}) = \rho + O(R^*(t_n^*))$. While it is certainly true that the bias goes to zero as $t_n \rightarrow 0^+$, the rate of convergence can be very slow. Recall from the rate of convergence comment in Subsection 2.2 that one possibility is $R^*(t_n^*) = -1/\ln t_n^*$, which would require an extremely large sample size to obtain a negligible (but still present) bias.

It is popular to form asymptotic confidence intervals for the tail exponent estimate ρ using the asymptotic normality property. However, when estimating the standard error of $\hat{\rho}$, replacing the true quantities with their estimates adds a bias to the estimated standard errors since the parameter estimates are biased. Therefore, confidence intervals for ρ are invalid except for very large sample sizes when the bias is sufficiently small.

6.2.2 The Case of $\rho = 0$. Many common parametric probability models are classified as medium tailed. For example, the Normal, Exponential, Weibull, Logistic, Lognormal, and Gumbel (or type I Extreme Value) distributions have tail exponent $\rho = 0$. However, the case $\rho = 0$ has not been discussed in the theoretical discussion of the proposed tail estimates.

This is predominantly due to the widely different types of tail behavior displayed by probability models in this class. For example, both the Normal and

Lognormal have $\rho = 0$, but from Figure 5 it is clear that the two distribution require very different tail estimates.

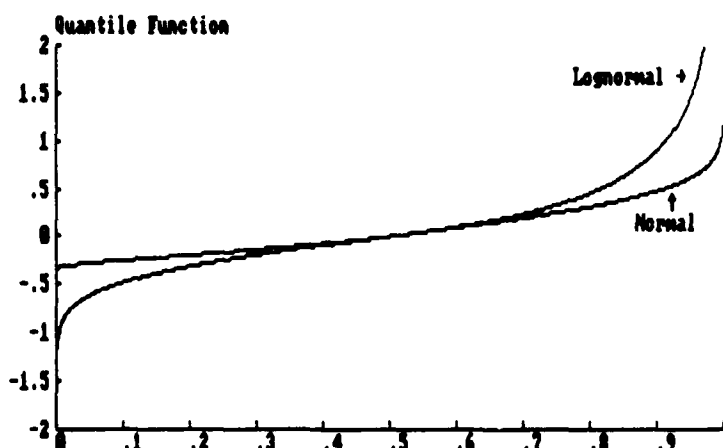


FIG. 5. Graph of the quantile function for the Normal and Lognormal distributions. Both these distributions have $\rho = 0$, but the tail from .75 to .99 should obviously be modeled differently.

The complication to the medium tailed class is that *ultimately*, the tail exponent is zero. That is, far enough out in the tail, all the distributions in the medium tailed class display the same tail behavior. In an attempt to differentiate between the tail behavior before this ultimate convergence, Schuster (1984) proposes the following subclasses:

- (i) $\rho = 0$, $L(u) \rightarrow \infty$ as $u \rightarrow 0^+ \Rightarrow$ *medium-short tails*;
- (ii) $\rho = 0$, $0 < \lim_{u \rightarrow 0^+} L(u) < \infty \Rightarrow$ *medium-medium tails*;
- (iii) $\rho = 0$, $L(u) \rightarrow 0$ as $u \rightarrow 0^+ \Rightarrow$ *medium-long tails*;

When the range of interest is not the extreme end of the tail, a *penultimate* model may better fit the observed tail. The ideas of ultimate versus penultimate behavior have long been recognized in extreme value theory. It is not surprising that these same issues occur in estimating tail behavior since the justification for the tail behavior model $fQ(1 - u) = u^{\rho+1}L(u)$ follows from extreme value theory.

The penultimate modeling of tail behavior is to allow a nonzero value for

ρ which more closely approximates the observed tail on the region of interest. This idea adds new problems to the identifiability of the true tail exponent ρ , but greatly improves the tail estimates.

6.2.3 Summary of Parameter Interpretation. Because the parameters are nonidentifiable and a value other than the true tail exponent may fit better to the observed tail, there are serious complications in interpreting the parameter estimates.

Therefore, one must choose which aspect of the tail is of interest and sample accordingly. If the tail exponent ρ is the focus of the analysis, an extremely large sample is required to make the bias negligible and force an extreme threshold value giving the ultimate convergence modeling.

6.3 Effect Of Bias In Parameter Estimates On Tail Estimates

If the tail estimates are of interest, the bias in the parameter estimates is less likely to seriously alter the tail estimates. The estimators using the GPD modeling and both Hall's estimators and Hill's estimators are an interesting special case when the bias is $O(R^*(t_n^*))$. The following theorem shows the bias in the tail estimates given in (2.3.5–2.3.7) is of the same rate in this special case.

THEOREM 6.3.1. Suppose that $fQ(1-u) = u^{\rho+1}L(u)$, where $\rho \in \mathbb{R}$, $\rho \neq 0$ and

$$\left| \frac{L(t)}{L(tu)} - 1 \right| \leq A(u)R(t)$$

for some positive measurable functions $A(u)$ and $R(t)$ where $\lim_{t \rightarrow 0+} R(t) = 0$. Let $\{T_n\}$ be a sequence of thresholds on $(Q(0), Q(1))$ such that $nt_n^* \rightarrow \infty$ and $t_n^* \rightarrow 0$ as $n \rightarrow \infty$, where $t_n = 1 - F(T_n)$. Further, let $a = a(t_n^*)$ be the scalar value of a function $a(\cdot)$ satisfying $a(t_n^*)hQ(1 - t_n^*) = 1 + O(R_1(t_n^*))$ for some positive measurable function $R_1(t)$ with $\lim_{t \rightarrow 0+} R_1(t) = 0$. Let $(\hat{\rho}_n, \hat{a}_n)$ denote the parameter estimates. Suppose that conditional on T_n ,

$$\begin{bmatrix} \hat{\rho}_n \\ \hat{a}_n \end{bmatrix} \text{ is AN } \left(\begin{bmatrix} \rho + O(R^*(t_n^*)) \\ a + O(R^*(t_n^*)) \end{bmatrix}, (nt_n^*)^{-1} \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} \right)$$

as $nt_n^* \rightarrow \infty$ for scalars v_{ij} such that the covariance matrix is positive definite.

(a) For $1 - t_n^* < u < 1$,

$$Q^*(u) \quad \text{is} \quad \text{AN} \left(Q(u) + \epsilon_Q^*(t_n^*, u), (nt_n^*)^{-1} \cdot \sigma_Q^2(u) \right)$$

where $|\epsilon_Q^*(t_n^*, u)| \leq A^{**}(u)R^*(t_n^*)$ for some positive measurable function $A^{**}(u)$, and $\sigma_Q^2(u)$ is given by (2.3.8).

(b) For $T_n < x < Q(1)$,

$$F^*(x) \quad \text{is} \quad \text{AN} \left(F(x) + \epsilon_F^*(t_n^*, x), (t_n^*/n) \cdot \sigma_F^2(x) \right)$$

where $|\epsilon_F^*(t_n^*, x)| \leq A^{**}(x) \cdot t_n^* R^*(t_n^*)$ for some positive measurable function $A^{**}(u)$, and $\sigma_F^2(x)$ is given by (2.3.9).

(c) For $T_n < x < Q(1)$,

$$f^*(x) \quad \text{is} \quad \text{AN} \left(f(x) + \epsilon_f^*(t_n^*, x), (t_n^*/n) \cdot \sigma_f^2(x) \right)$$

where $|\epsilon_f^*(t_n^*, x)| \leq A^{**}(x) \cdot t_n^* R^*(t_n^*)$ for some positive measurable function $A^{**}(u)$, and $\sigma_f^2(x)$ is given by (2.3.10).

The important result from this theorem is that the bias in the tail estimates is of the same order as the bias in the parameter estimates. The bias is negligible for t_n^* sufficiently small, but the idea of 'sending t_n^* to zero' is unacceptable in applications since this effectively removes the desire to estimate the tail. Therefore, some amount of bias must be accepted due to the very nature of the problem.

7. THRESHOLD SELECTION

In order to reduce the bias in the parameter estimates proposed in this work, the threshold percentile t_n should be chosen so that $\lim_{n \rightarrow \infty} t_n = 0$. However, the asymptotic results for these estimators require that $\lim_{n \rightarrow \infty} nt_n = \infty$. Therefore, a trade off between bias and variance must be made when choosing the threshold value.

Further, the concept of an 'optimal' threshold percentile can be viewed one of two ways. If the tail behavior parameters are the focus of the analysis, then a threshold should be chosen which minimizes some criterion function for the parameter estimate (such as MSE). However, if the tail estimates are the focus of the analysis, then the threshold should be chosen which best fits the observed tail.

7.1 Optimal Thresholds Based On The Parameters

Csörgő, Horváth, and Révész (1987) investigate the existence of optimal sequences for the threshold for any estimate of the tail exponent. They conclude that since a sequence for the threshold based on some optimality criterion depends on the unknown tail exponent and slowly varying function, it is useless in practice.

In an attempt to avoid this dependency on unknown quantities, Hall and Welsh (1985) considered Hill's estimator and proposed a parametric model for the slowly varying function. They propose an optimal threshold sequence $t_{opt} = C_0(\gamma, \Delta) \cdot n^{-1/2\gamma}$, where $C_0(\gamma, \Delta)$ is selected according to some criterion such as MSE, under the assumption $L(u) = (1/\rho\Delta)[1 - \Theta u^\gamma + o(u^\gamma)]$, for $\gamma > 0$ and $\Delta > 0$. Hall and Welsh then propose estimates for γ and Δ which are used to estimate the optimal threshold sequence. This procedure is highly dependent on the quality of the parameterization for $L(u)$, and the associated parameter estimates performed quite poorly in their simulation study.

7.2 Optimal Thresholds Based On The Tail

When the focus is on the tails, the 'best' threshold is defined as the one that provides a tail estimate 'closest' to the observed tail. This leads to a data analytic threshold selection procedure where the observed tail dictates the threshold value.

The following algorithm is proposed for choosing a threshold percentile t_n which best approximates the sample.

- For $t = 1/n, 2/n, \dots, (n-1)/n$:
 1. Compute $F_t^{\wedge}(u)$ based on the exceedences $X_j - Q^{\sim}(1-t)$ given $X_j > Q^{\sim}(1-t)$.
 2. Compute

$$d_t = \sup_{Q^{\sim}(1-t) < x < Q^{\sim}(1)} |F^{\sim}(x) - F_t^{\wedge}(x)|.$$

- Choose $t_{\text{opt}} = t$, where t is the smallest solution to $d = \min_t d_t$.

This algorithm chooses the threshold percentile which minimizes the distance between the estimated distribution function and the sample distribution function over the tail values. It is an attempt to satisfy the intuitive behavior of the threshold percentile, where t_n is small enough to make the bias small and nt_n is large enough so that the tail estimates are based on a large number of observations. The properties of this threshold selection procedure warrant further study.

8. MOTIVATING EXAMPLE

The importance of estimating the frequency and magnitude of flood discharges in rivers is obvious. In most instances, the 'average' flood discharge is of little interest. Instead, the focus of inference is on floods with small probabilities of occurrence. This naturally then leads to investigation of the tail of the underlying distribution.

Frequently estimated quantities in hydrology are the 'hundred year flood' or 'thousand year flood,' i.e. the flood discharge such that a flood of this magnitude (or greater) will occur once in 100 years or 1000 years (on average). These are convenient synonyms for $Q(.99)$ and $Q(.999)$, respectively. Therefore, estimating the tail of the underlying quantile function is the main objective of the analysis.

Two data sets from hydrology have been obtained from the flood frequency analyses discussed in Pericchi and Rodríguez-Iturbe (1985). They consider two rivers in the United States that are of importance for the regions in which they are located and whose floods have been the object of previous calculations for engineering works. The first data set is taken from Benjamin and Cornell (1970) and contains 59 years of annual floods (1902–1960) for the Feather River at Oroville, California. The Feather River annual floods are given in Table 6 along with descriptive statistics for location, scale, and tail behavior.

The second data set is taken from Wood, Rodríguez-Iturbe, and Schaake (1974) and contains 37 years of annual floods (1929–1965) for the Blackstone River at Woonsocket, Rhode Island. The Blackstone River annual floods are given in Table 7 along with descriptive statistics for location, scale, and tail behavior.

The IQ box plots for the Feather River data and the Blackstone River data are given in Figures 6a and 6b, respectively. Both distributions are skewed and heavy tailed, as is expected in flood data. Large values in the annual flood sequence are often interpreted as outliers since they are difficult to model, yet are expected due to nature and should be deleted only in special circumstances.

The most important but also the most difficult problem in analysis is the

TABLE 6

Annual floods (1902-1960) for the Feather River at Oroville, California taken from Benjamin and Cornell (1970). Descriptive statistics for location, scale, and right tail behavior are also given. Notice that the underlying probability model is skewed and heavy tailed.

Year	Flood Discharge (ft ³ /sec)	Year	Flood Discharge (ft ³ /sec)
1902	42,000	1932	22,600
1903	102,000	1933	8,860
1904	118,000	1934	20,300
1905	81,000	1935	58,600
1906	128,000	1936	85,400
1907	230,000	1937	19,200
1908	16,300	1938	185,000
1909	140,000	1939	8,080
1910	31,000	1940	152,000
1911	75,400	1941	84,200
1912	16,400	1942	110,000
1913	16,800	1943	108,000
1914	122,000	1944	24,900
1915	81,400	1945	60,100
1916	42,400	1946	54,400
1917	80,400	1947	45,600
1918	28,200	1948	36,700
1919	65,900	1949	16,800
1920	23,400	1950	46,400
1921	62,300	1951	92,100
1922	36,400	1952	59,200
1923	22,400	1953	113,000
1924	42,400	1954	54,800
1925	64,300	1955	13,000
1926	55,700	1956	203,000
1927	94,000	1957	83,100
1928	185,000	1958	102,000
1929	14,000	1959	34,500
1930	80,100	1960	135,000
1931	11,600		

Measures of Center

Median: 58,600.0
Mean: 70,265.1

Measures of Dispersion

Interquartile Range 70,600.0
Standard Deviation 52,023.5

Measures of Right Tail Behavior

$Q\Gamma(.90) = .5411$ $Q\Gamma(.95) = .8952$ $Q\Gamma(.99) = 1.0227$

TABLE 7

Annual floods (1929–1965) for the Blackstone River at Woonsocket, Rhode Island taken from Wood, Rodríguez-Iturbe, and Schaake (1974). Descriptive statistics for location, scale, and right tail behavior are also given. Notice that the underlying probability model is skewed and heavy tailed.

Year	Flood Discharge (ft ³ /sec)	Year	Flood Discharge (ft ³ /sec)
1929	4,570	1948	5,810
1930	1,970	1949	2,030
1931	8,220	1950	3,620
1932	4,530	1951	4,920
1933	5,780	1952	4,090
1934	6,560	1953	5,570
1935	7,500	1954	9,400
1936	15,000	1955	32,900
1937	6,340	1956	8,710
1938	15,100	1957	3,850
1939	3,840	1958	4,970
1940	5,860	1959	5,398
1941	4,480	1960	4,780
1942	5,330	1961	4,020
1943	5,310	1962	5,790
1944	3,830	1963	4,510
1945	3,410	1964	5,520
1946	3,830	1965	5,300
1947	3,150		

Measures of Center

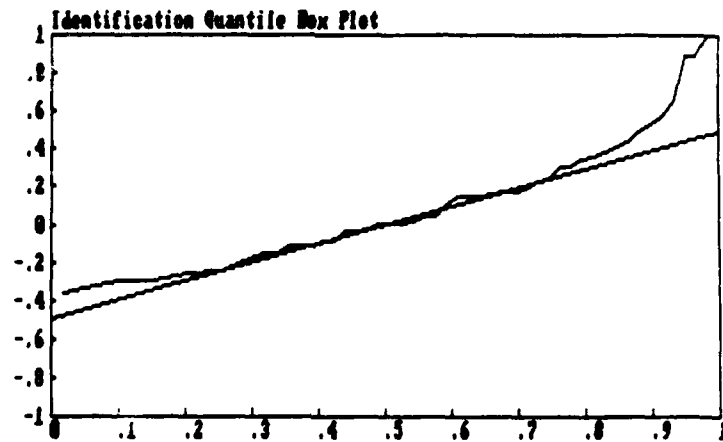
Median: 4,970.0
Mean: 6,372.9

Measures of Dispersion

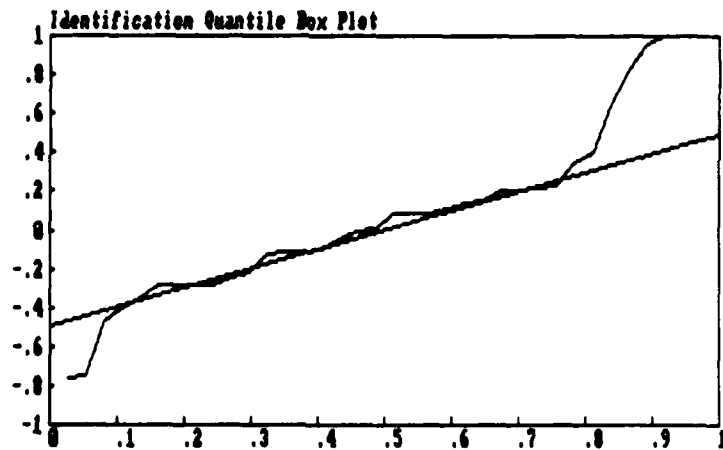
Interquartile Range 1,960.0
Standard Deviation 5,276.7

Measures of Right Tail Behavior

$Q\Gamma(.90) = .9541$ $Q\Gamma(.95) = 2.5587$ $Q\Gamma(.99) = 2.5842$



(A)



(B)

FIG. 6. Identification quantile box plots for the data used as examples. The identification quantile box plot is a truncated plot of the sample identification quantile function with an overlaid reference line to help in assessing the shape of the underlying distribution. Figure 6a describes the Feather River annual floods for the years 1902–1960 measured at Oroville, California. Figure 6b describes the Blackstone River annual floods for the years 1929–1965 measured at Woonsocket, Rhode Island. For each data set, it appears the underlying distribution is skewed and heavy tailed as is expected in flood data. The largest observations of the Blackstone River data can be interpreted as outliers, but large observations in flood data are certainly expected due to nature so deleting data points is acceptable only under special circumstances.

selection of an appropriate probability model for the annual floods. The implicit assumption of the analysis is that the probability model fit to the observed annual floods is valid beyond the observed range of values. Three common parametric probability models used in flood frequency analysis are the Gumbel (also referred to as the Type I Extreme Value) distribution, the Lognormal distribution, and the Pearson Type III (also referred to as the three parameter gamma) distribution.

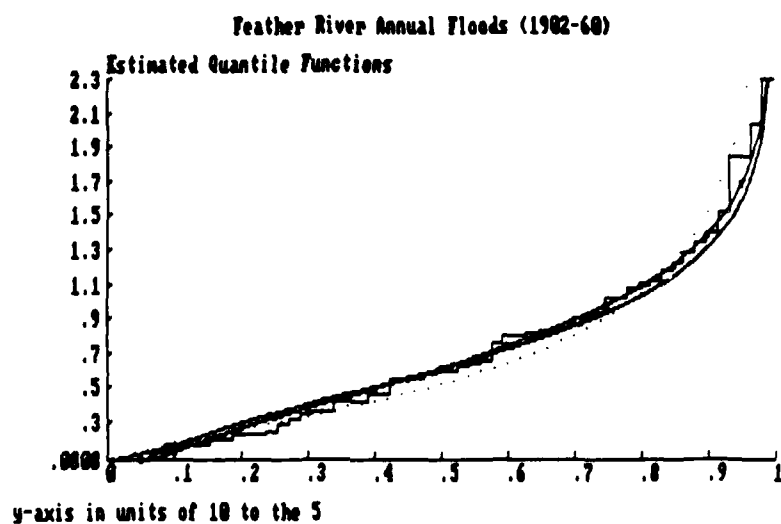
The Gumbel can be theoretically motivated since the annual flood can be considered as the maximum of many independent and identically distributed random variables from any population classified as medium tailed (*e.g.* normal, exponential, Weibull, lognormal, logistic). The Lognormal and Pearson type III are convenient in that they have a small number of parameters and are flexible in fitting the data.

Each of these probability models has been fit to the Feather River data and the Blackstone River, and interposed over the sample quantile function $Q^*(u)$ in Figures 7a and 7b, respectively. Notice that all three models appear reasonable for describing the annual floods. In fact, using the Kolmogorov-Smirnov goodness of fit test, all fail to reject at $\alpha = .05$.

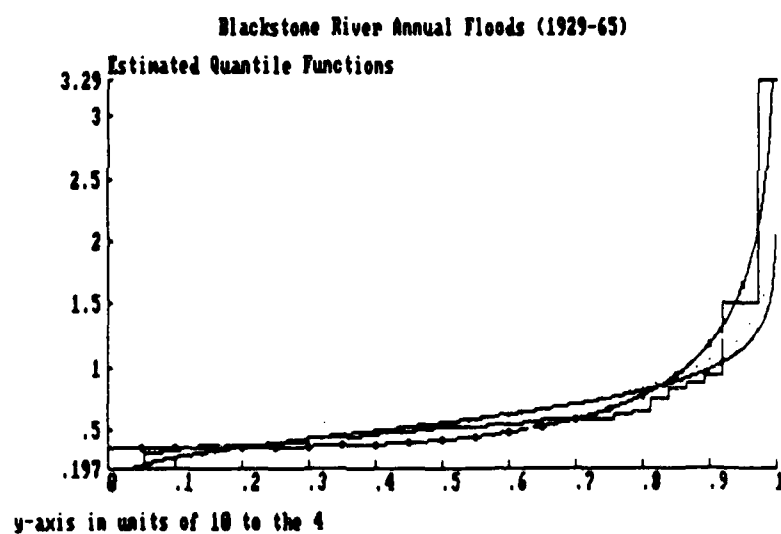
Since the tail of the quantile function is the focus of the analysis, consider Figures 8a and 8b which graph these same functions on the upper quantile. For the upper tail, these 'acceptable' models provide very different estimates. This is not surprising since the estimated probability model was fit using the entire sample and the model will best fit the center where the majority of the observations lie.

In order to remove the influence of observations at the center, estimators based on the exceedences of a threshold are appropriate. Further, a generally applicable estimator, free of the restrictions imposed by a parametric family, will allow the data to dictate the tail estimate. Therefore, the tail estimates proposed in this work will provide useful tools for improving the analysis of annual flood data.

Before computing the estimates proposed in this work for the Feather River data and the Blackstone River data, the observations have been centered and

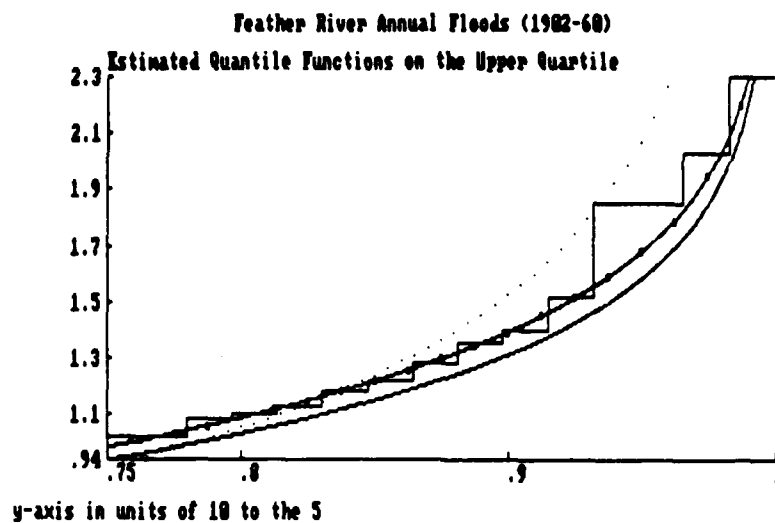


(A)

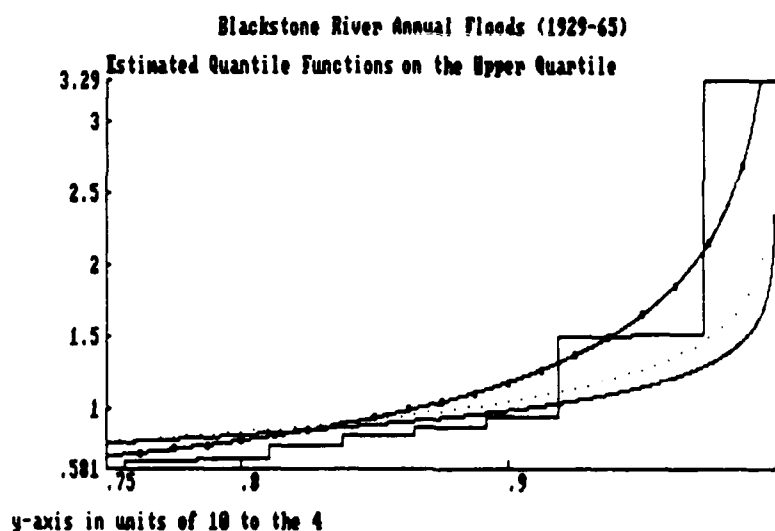


(B)

FIG. 7. Graph of the estimated Gumbel model (solid line), Lognormal model (dotted line), and Pearson Type III model (solid line with blocks) overlaid on the sample quantile function (step function) for the data used as examples. Figure 7a is for the Feather River data and Figure 7b is for the Blackstone River data. All three models appear reasonable for describing the annual floods, and all fail to reject the Kolmogorov-Smirnov goodness of fit test at $\alpha = .05$.



(A)



(B)

FIG. 8. Graph of the estimated Gumbel model (solid line), Lognormal model (dotted line), and Pearson Type III model (solid line with blocks) overlaid on the sample quantile function (step function) on the upper quartile for the data used as examples. Figure 8a is for the Feather River data and Figure 8b is for the Blackstone River data. Notice that the three probability models provide very different tail estimates. This is due to the use of the entire sample in the estimation, forcing the model to fit better at the center of the distribution where the majority of the observations lie.

scaled by subtracting the median and dividing by twice the interquartile range. The parameter estimates are not location-scale invariant, and a different standardization will result in different parameter estimates. However, this is not a serious issue since the tail estimates are the focus of the analysis and can be obtained in the original units by correcting for the standardization.

8.1 Tail Estimates For The Feather River

The parameter estimates for the GPD and GEV modeling of the exceedences are given in Table 8. Since it appears reasonable from the IQ box plot of the Feather River Data to classify the underlying probability model as short tailed, Hall's parameter estimates are also appropriate and tabled. A further implication of this classification is the existence of a finite upper bound $Q(1)$. This does not imply that a 'maximum flood level' exists, but that the best approximating probability model has such an upper bound. A graphical comparison of the different parameter estimators is given in Figure 9.

8.2 Tail Estimates For The Blackstone River

The parameter estimates for the GPD and GEV modeling of the exceedences are given in Table 9. Since it appears reasonable from the IQ box plot of the Blackstone River Data to classify the underlying probability model as long tailed, Hill's parameter estimates are also appropriate and tabled. It is interesting to note that the threshold selection algorithm chose the same threshold percentile for each of the proposed tail estimators. However, even with the same threshold percentile the parameter estimates are different. A graphical comparison of the different parameter estimators is given in Figure 10.

TABLE 8

Table containing the optimal threshold percentile and parameter estimates for the proposed tail estimates based on the exceedences of a threshold for the Feather River annual floods. Parameters are estimated using the GPD modeling, the GEV modeling, and Hall's estimate for an underlying short tailed distribution. The classification of the underlying distribution as short tailed follows from the IQ box plot. Notice that since a short tailed probability model appears reasonable, a finite upper bound $Q(1)$ is estimated by the GPD estimate and Hall's estimate. The GEV estimate has $Q(1) = \infty$ since $\hat{\rho} > 0$.

	Threshold Percentile t	$\hat{\rho}$	\hat{a}	$Q(1)$
GPD	20/59 \approx .339	-.259	.426	290,571 ft ³ /sec
GEV	23/59 \approx .390	.021	.254	
Hall's Est.	24/59 \approx .407	-.106	.341	528,615 ft ³ /sec

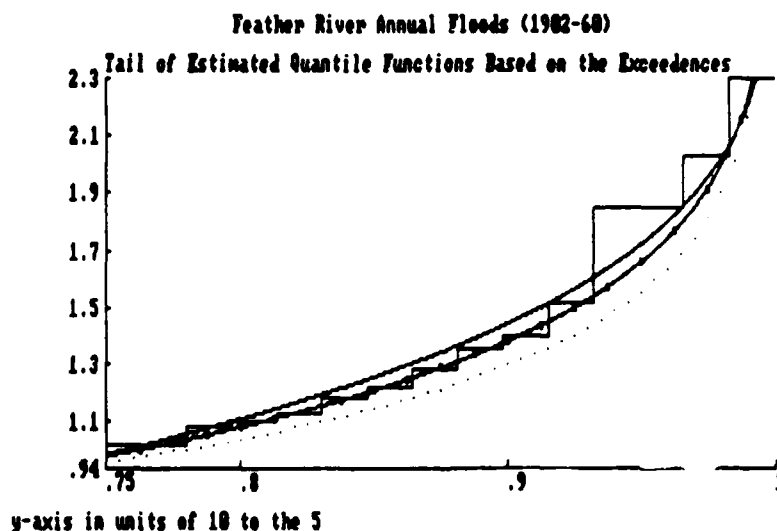


FIG. 9. Graph of the proposed tail estimates based on the exceedences of a threshold for the Feather River annual floods on the upper quartile. The tail estimate based on the GPD modeling (solid line), GEV modeling (dotted line), and using Hall's estimate for an underlying short tailed probability model (solid line with blocks) are overlaid on the sample quantile function (step function) for comparison.

TABLE 9

Table containing the optimal threshold percentile and parameter estimates for the proposed tail estimates based on the exceedences of a threshold for the Blackstone River annual floods. Parameters are estimated using the GPD modeling, the GEV modeling, and Hill's estimate for an underlying long tailed distribution. The classification of the underlying distribution as long tailed follows from the IQ box plot.

	Threshold Percentile t	$\hat{\rho}$	\hat{a}
GPD	$18/37 \approx .486$	1.100	.218
GEV	$18/37 \approx .486$	1.076	.131
Hill's Est.	$18/37 \approx .486$	1.602	.135

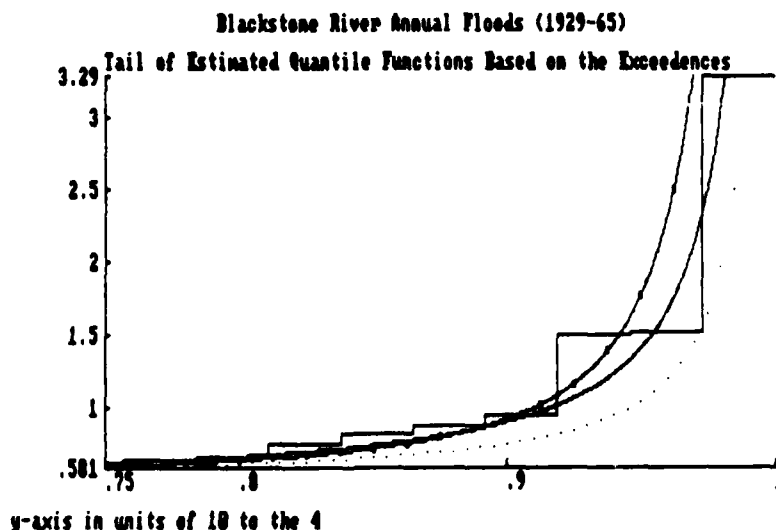


FIG. 10. Graph of the proposed tail estimates based on the exceedences of a threshold for the Blackstone River annual floods on the upper quartile. The tail estimate based on the GPD modeling (solid line), GEV modeling (dotted line), and using Hill's estimate for an underlying long tailed probability model (solid line with blocks) are overlaid on the sample quantile function (step function) for comparison.

9. CONCLUDING REMARKS

Tail estimates of the underlying probability model are of interest in many applications. In addition, the tail behavior of a probability model dictates many theoretical properties with important implications for probability modeling. Therefore, generally applicable tail estimates can serve as valuable diagnostic tools in fitting probability models to data.

In this work, tail estimates have been proposed which:

- use only the observations in the tail, and
- are generally applicable, making minimal assumptions on the underlying probability model.

Two distinct approaches in this format are unified by modeling the density-quantile function as a regularly varying function and representing the quantile function for the conditional distribution of the exceedences of a threshold as the sum of a parametric function and an analytic error function. The quantile representation for the exceedences is the key to

- (1) forming a parametric model for the tail of the underlying probability model;
- (2) motivating methods for obtaining parameter estimates; and
- (3) deriving the asymptotic properties of the proposed parameter estimates.

Parameter estimates may be obtained using a Generalized Pareto Distribution (GPD) or a Generalized Extreme Value Distribution (GEV) modeling of the exceedences. Assuming the underlying distribution can be correctly classified as either short tailed or long tailed, other estimates can be formed.

The unified approach allows for comparison of these different estimators. All are shown to be biased, and no global statements can be made regarding an 'optimal' estimator.

Much of the previous work on estimating tail behavior has focused solely on the problem of parameter estimation. However, the parameters are shown to be nonidentifiable and their estimators will always contain a bias which may be non-negligible. In order to estimate the parameters with a reasonable amount of precision, extremely large sample sizes are required.

In this work, the focus has been on obtaining tail estimates. In this scenario, the bias in the parameter estimates causes a bias in the tail estimate of the same order. Bias reduction is made at the cost of inflated variance. Therefore, some compromise must be made.

To demonstrate the tail estimators proposed in this work, two sequences of annual floods were selected. The problem of probability modeling is an important issue in hydrology since the estimate of the tail is highly dependent on the model and there is little empirical evidence one can produce to support a given model. Thus, a generally applicable approach to tail estimation where the data dictates the form of the model is a useful diagnostic tool for further probability modeling.

REFERENCES

- BENJAMIN, J. R. and CORNELL, C. A. (1970). *Probability, Statistics and Decision for Civil Engineers*. McGraw-Hill, New York.
- BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987). *Regular Variation*. Cambridge Univ. Press, London.
- BLACKMAN, D. and GRAFF, J. (1978). The analysis of annual extreme sea levels at certain ports in southern England. *Proc. Instn. Civ. Engrs.*, Part 2 **65** 339-357.
- BOX, G. E. P. and COX, D. R. (1964). An analysis of transformations. *J. Roy. Statist. Soc. Ser. B* **26** 211-252.
- CSÖRGŐ, M., HORVÁTH, L. and RÉVÉSZ, P. (1987). On the optimality of estimating the tail index and a naive estimator. *Austral. J. Statist.* **29** 166-178.
- CSÖRGŐ, M. and RÉVÉSZ, P. (1975). Some notes on the empirical distribution function and the quantile process. In *Limit Theorems of Probability* 59-71. North Holland, Amsterdam.
- CSÖRGŐ, M. and RÉVÉSZ, P. (1978). Strong approximations of the quantile process. *Ann. Statist.* **6** 882-894.
- DAVIS, R. and RESNICK, S. (1984). Tail estimates motivated by extreme value theory. *Ann. Statist.* **12** 1467-1487.
- DAVISON, A. C. (1984). Modelling excesses over high thresholds, with an application. In *Statistical Extremes and Applications* (J. Tiago de Oliveira, ed.) 461-482. Reidel, Dordrecht.
- DUMOUCHEL, W. (1983). Estimating the stable index α in order to measure tail thickness: A Critique. *Ann. Statist.* **11** 1019-1036.

- FISHER, R. A. (1948). *Statistical Methods for Research Workers*. Oliver and Boyd, London.
- FISHER, R. A. and TIPPETT, L. H. C. (1928). Limiting forms of the frequency of the largest or smallest member of a sample. *Proc. Camb. Phil. Soc.* **24** 180-190.
- GOLDIE, C. M. and SMITH, R. L. (1987). Slow variation with remainder: A survey of the theory and its applications. *Quart. J. Math. Oxford. Ser. (2)* **38** 45-71.
- HALL, P. (1982). On estimating the endpoint of a distribution. *Ann. Statist.* **10** 556-568.
- HALL, P. and WELSH, A. H. (1985). Adaptive estimates of parameters of regular variation. *Ann. Statist.* **13** 330-341.
- HILL, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3** 1163-1174.
- HOSKING, J. R. M. (1985). Maximum-likelihood estimation of the parameters of the Generalized Extreme Value Distribution. *Appl. Statist. (Ser. C)* **34** 301-310.
- HOSKING, J. R. M. and WALLIS, J. R. (1987). Parameter and quantile estimation for the Generalized Pareto Distribution. *Technometrics* **29** 339-349.
- JENKINSON, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quart. J. Roy. Met. Soc.* **81** 158-171.
- JOE, H. (1987). Estimation of quantiles of the maximum of N observations. *Biometrika* **74** 347-354.
- MASON, D. M. (1982). Laws of large numbers for sums of extreme values. *Ann. Probab.* **10** 754-764.

- NERC (1975). *Flood Studies Report, 1*. Natural Environment Research Council, London.
- PARZEN, E. (1961a). An approach to time series analysis. *Ann. Math. Statist.* **32** 951-989.
- PARZEN, E. (1961b). Regression analysis of continuous parameter time series. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 469-489. Univ. California Press, Berkeley.
- PARZEN, E. (1967). *Time Series Analysis Papers*. Holden-Day, San Francisco.
- PARZEN, E. (1979). Nonparametric statistical data modeling. *J. Amer. Statist. Assoc.* **74** 105-131.
- PARZEN, E. (1983). Informative quantile functions and identification of probability distribution types. *Proc. 29th Army Conf. Design of Experiments*, ARO Report 84-2, 97-107.
- PERICCHI, L. R. and RODRÍGUEZ-ITURBE, I. (1985). On the statistical analysis of floods. In *A Celebration of Statistics*, (A. C. Atkinson and S. E. Fienberg, eds.) 511-541. Springer-Verlag, New York.
- PICKANDS, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3** 119-131.
- SCHUSTER, E. F. (1984). Classification of probability laws by tail behavior. *J. Amer. Statist. Assoc.* **79** 936-939.
- SMITH, J. A. (1986). Estimating the upper tail of flood frequency distributions. Technical Report MS-R8607, Dept. Math. Statist., Centrum voor Wiskunde en Informatica, Amsterdam.

- SMITH, R. L. (1984). Threshold methods for sample extremes. In *Statistical Extremes and Applications* (J. Tiago de Oliveira, ed.) 621-638. Reidel, Dordrecht.
- SMITH, R. L. (1987). Estimating tails of probability distributions. *Ann. Statist.* 15 1174-1207.
- SMITH, R. L. and WEISSMAN, I. (1985). Maximum likelihood estimation of the lower tail of a probability distribution. *J. Roy. Statist. Soc. Ser. B* 47 285-298.
- WOOD, E. F., RODRÍGUEZ-ITURBE, I., and SCHAAKE, J. C. (1974). The methodology of Bayesian inference and decision making applied to extreme hydrologic events. Report No. 178, Ralph M. Parsons Lab., M. I. T., Cambridge.

APPENDIX A

COMMON PARAMETRIC PROBABILITY MODELS

This appendix contains the identification standardized versions of many common parametric random variables. The identification transformation is useful in comparing different types of tail behavior since the corresponding quantile function equals zero and has slope approximately one at $u = .5$.

The identification transformation of a random variable Y with distribution function $F(y)$, density function $f(y)$, and quantile function $Q(u)$ is simply a location - scale standardization. Let $\mu = Q(.5)$, the median, and $\sigma = 2[Q(.75) - Q(.25)]$, the quartile deviation. Then make the transformation $ZQI = (Y - \mu)/\sigma$, which results in the identification distribution function $FI(z) = F(\mu + \sigma z)$, identification density function $fI(z) = \sigma f(\mu + \sigma z)$, and identification quantile function $QI(u) = (Q(u) - \mu)/\sigma$.

This appendix defines the distribution function, density function, quantile function, and density-quantile function for the identification standardized versions of some common parametric probability models. Graphs of these functions are also given.

- The Uniform Distribution is given by

$$F(x) = \begin{cases} 0, & x \leq -.5 \\ x + .5, & -.5 < x < .5 \\ 1, & x \geq .5, \end{cases}$$

with quantile function

$$Q(u) = u - .5,$$

density function

$$f(x) = \begin{cases} 1, & -.5 < x < .5 \\ 0, & \text{otherwise,} \end{cases}$$

and density-quantile function

$$fQ(u) = 1.$$

These functions are plotted in Figure 11.

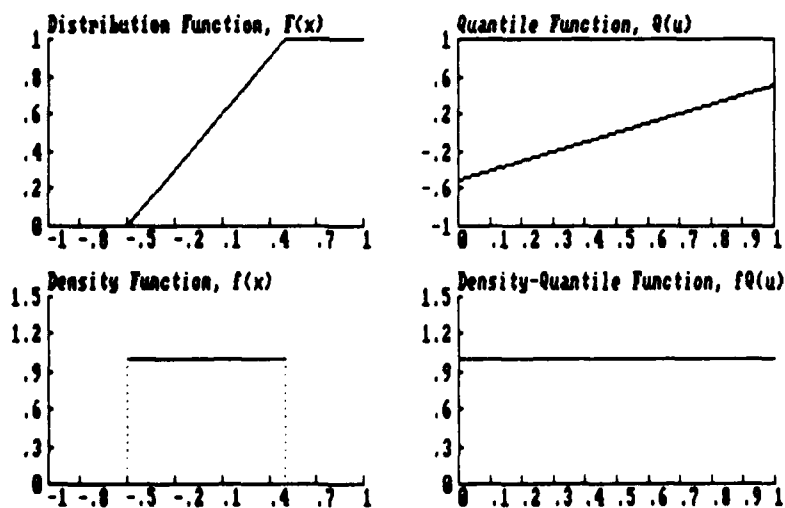


FIG. 11. Uniform Distribution.

- The Negative Exponential Distribution is given by

$$F(x) = \begin{cases} \frac{1}{2}e^{(2\ln 3)x}, & x < \ln 2/(2\ln 3) \\ 1, & x \geq \ln 2/(2\ln 3), \end{cases}$$

with quantile function

$$Q(u) = \frac{\ln 2u}{2\ln 3},$$

density function

$$f(x) = \begin{cases} \ln 3 e^{(2\ln 3)x}, & x < \ln 2/(2\ln 3) \\ 0, & x \geq \ln 2/(2\ln 3), \end{cases}$$

and density-quantile function

$$fQ(u) = (2\ln 3)u.$$

These functions are plotted in Figure 12.

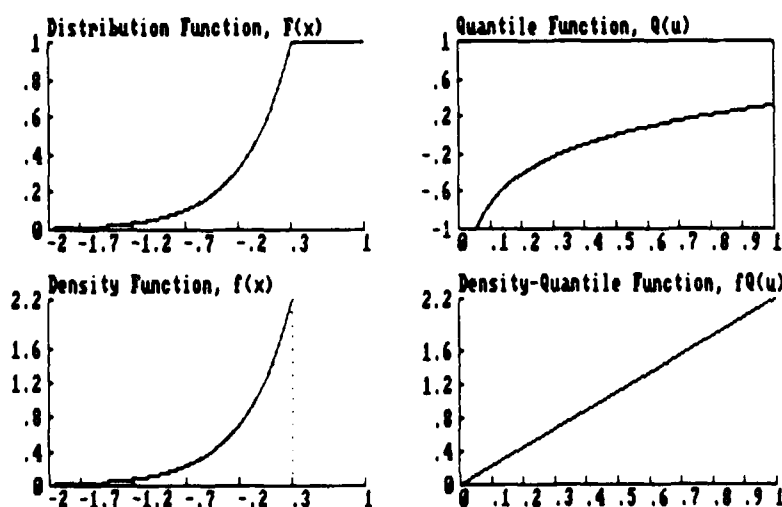


FIG. 12. Negative Exponential Distribution.

- The Negative Weibull(p) Distribution, $p > 0$, (which is also referred to as the Type III Extreme Value Distribution) is given by

$$F(x) = \begin{cases} \exp\{-(\ln 2)^{1/p} - \sigma x\}^p, & x < (\ln 2)^{1/p}/\sigma \\ 1, & x \geq (\ln 2)^{1/p}/\sigma, \end{cases}$$

with quantile function

$$Q(u) = \frac{(\ln 2)^{1/p} - (-\ln u)^{1/p}}{\sigma},$$

density function

$$f(x) = \begin{cases} \sigma p [(\ln 2)^{1/p} - \sigma x]^{p-1} \cdot \exp\{-(\ln 2)^{1/p} - \sigma x\}^p, & x < (\ln 2)^{1/p}/\sigma \\ 0, & x \geq (\ln 2)^{1/p}/\sigma, \end{cases}$$

and density-quantile function

$$fQ(u) = \sigma p u (-\ln u)^{-(1/p)+1},$$

where $\sigma = 2[(\ln 4)^{1/p} - (\ln 4 - \ln 3)^{1/p}]$. These functions are plotted in Figure 13.

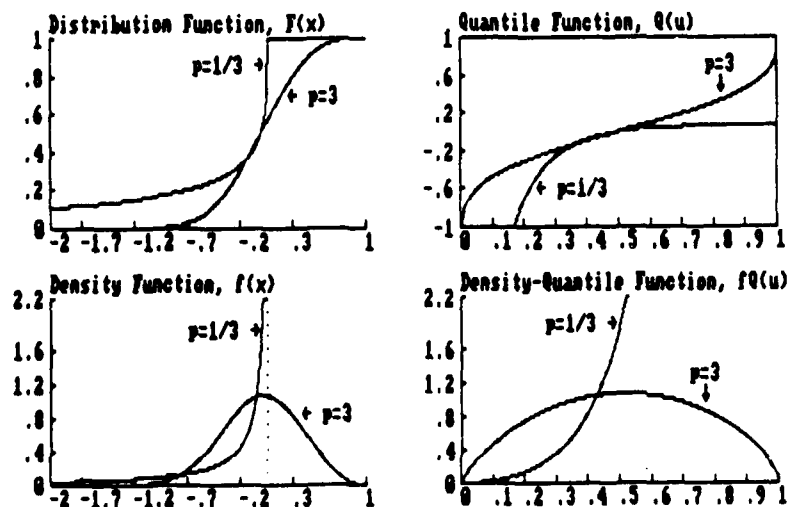


FIG. 13. Negative Weibull(p) Distribution.

- The Exponential Distribution is given by

$$F(x) = \begin{cases} 0, & x \leq -\ln 2 / (2 \ln 3) \\ 1 - \frac{1}{2} e^{-(2 \ln 3)x}, & x > -\ln 2 / (2 \ln 3), \end{cases}$$

with quantile function

$$Q(u) = \frac{-\ln 2(1-u)}{2 \ln 3},$$

density function

$$f(x) = \begin{cases} 0, & x \leq -\ln 2 / (2 \ln 3) \\ (\ln 3) e^{-(2 \ln 3)x}, & x > -\ln 2 / (2 \ln 3), \end{cases}$$

and density-quantile function

$$fQ(u) = (2 \ln 3)(1-u).$$

These functions are plotted in Figure 14.

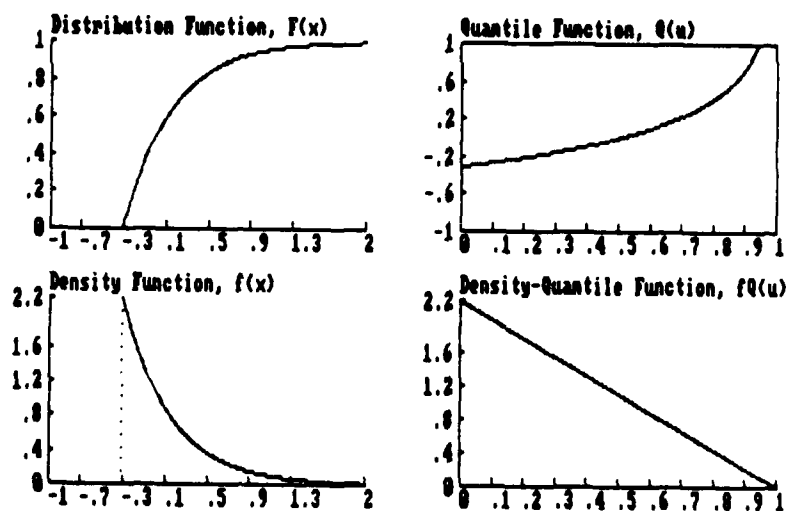


FIG. 14. Exponential Distribution.

- The Logistic Distribution is given by

$$F(x) = \frac{1}{1 + e^{-(4 \ln 3)x}},$$

with quantile function

$$Q(u) = \frac{\ln u - \ln(1 - u)}{4 \ln 3},$$

density function

$$f(x) = \frac{(4 \ln 3)e^{-(4 \ln 3)x}}{(1 + e^{-(4 \ln 3)x})^2},$$

and density-quantile function

$$fQ(u) = (4 \ln 3)u(1 - u).$$

These functions are plotted in Figure 15.

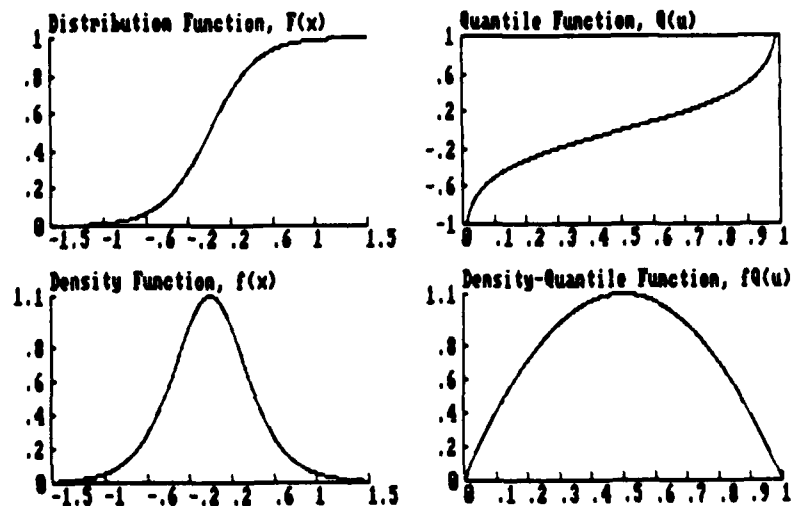


FIG. 15. Logistic Distribution.

- The Normal Distribution is given by

$$F(x) = \Phi(\sigma x),$$

with quantile function

$$Q(u) = \Phi^{-1}(u)/\sigma,$$

density function

$$f(x) = \sigma \phi(\sigma x),$$

and density-quantile function

$$\begin{aligned} fQ(u) &= \frac{\sigma}{\sqrt{2\pi}} \exp \left\{ -\frac{[\Phi^{-1}(u)]^2}{2} \right\} \\ &= \sigma \phi [\Phi^{-1}(u)], \end{aligned}$$

where $\Phi(x) = \int_{-\infty}^x \phi(t)dt$, $\phi(x) = (1/\sqrt{2\pi})e^{-x^2/2}$ and $\sigma = 2[\Phi^{-1}(.75) - \Phi^{-1}(.25)] \approx 2.6979$. These functions are plotted in Figure 16.

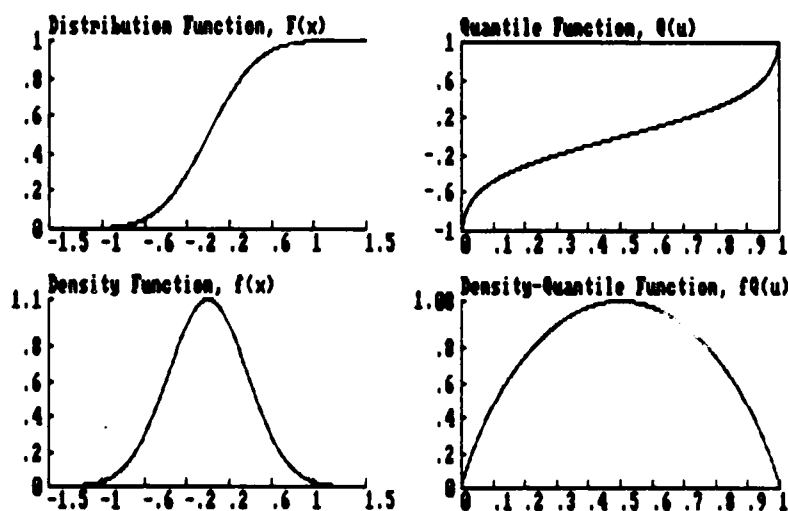


FIG. 16. Normal Distribution.

- The Weibull(p) Distribution, $p > 0$, is given by

$$F(x) = \begin{cases} 0, & x \leq -(\ln 2)^{1/p}/\sigma \\ 1 - \exp\{-[(\ln 2)^{1/p} + \sigma x]^p\}, & x > -(\ln 2)^{1/p}/\sigma, \end{cases}$$

with quantile function

$$Q(u) = \frac{[-\ln(1-u)]^{1/p} - (\ln 2)^{1/p}}{\sigma},$$

density function

$$f(x) = \begin{cases} 0, & x \leq -(\ln 2)^{1/p}/\sigma \\ \sigma p [(\ln 2)^{1/p} + \sigma x]^{p-1} \cdot \exp\{-[(\ln 2)^{1/p} + \sigma x]^p\}, & x > -(\ln 2)^{1/p}/\sigma, \end{cases}$$

and density-quantile function

$$fQ(u) = \sigma p (1-u) [-\ln(1-u)]^{-(1/p)+1},$$

where $\sigma = 2[(\ln 4)^{1/p} - (\ln 4 - \ln 3)^{1/p}]$. These functions are plotted in Figure 17.

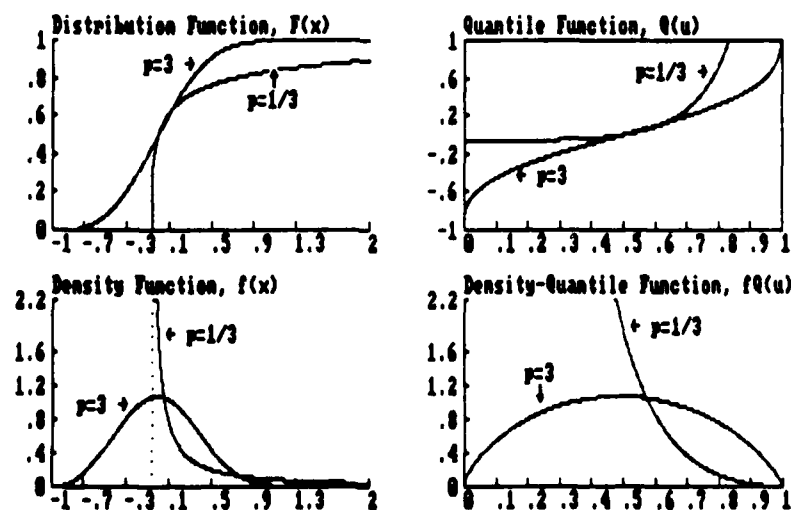


FIG. 17. Weibull(p) Distribution.

- The Lognormal Distribution is given by

$$F(x) = \Phi[\ln(\sigma x + 1)],$$

with quantile function

$$Q(u) = \frac{e^{\Phi^{-1}(u)} - 1}{\sigma},$$

density function

$$f(x) = \frac{\sigma}{\sigma x + 1} \phi[\ln(\sigma x + 1)],$$

and density-quantile function

$$fQ(u) = \sigma \phi[\Phi^{-1}(u)] \cdot e^{-\Phi^{-1}(u)},$$

where $\sigma = 2[e^{\Phi^{-1}(.75)} - e^{\Phi^{-1}(.25)}] \approx 2.9072$. These functions are plotted in Figure 18.

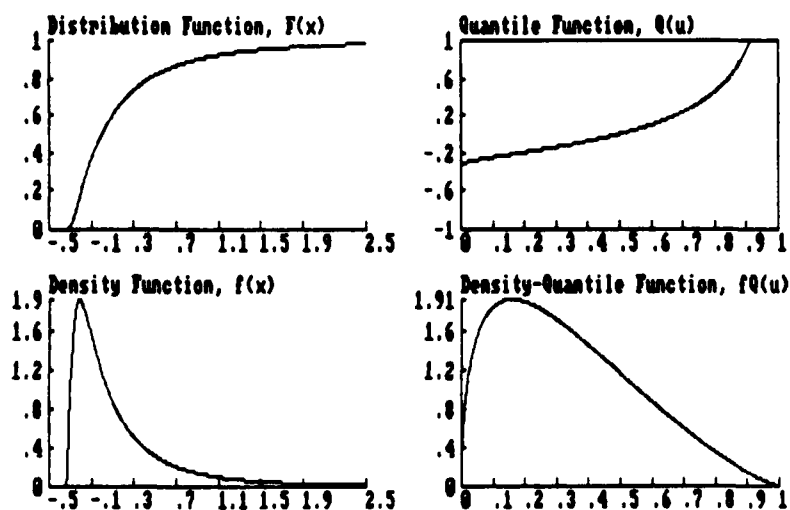


FIG. 18. Lognormal Distribution.

- The Cauchy Distribution is given by

$$F(x) = \frac{1}{2} + \frac{\tan^{-1} 4x}{\pi},$$

with quantile function

$$Q(u) = \frac{1}{4} \tan \pi \left(u - \frac{1}{2} \right),$$

density function

$$f(x) = \frac{4}{\pi(1 + 16x^2)},$$

and density-quantile function

$$fQ(u) = \frac{4}{\pi} \sin^2 \pi u.$$

These functions are plotted in Figure 19.

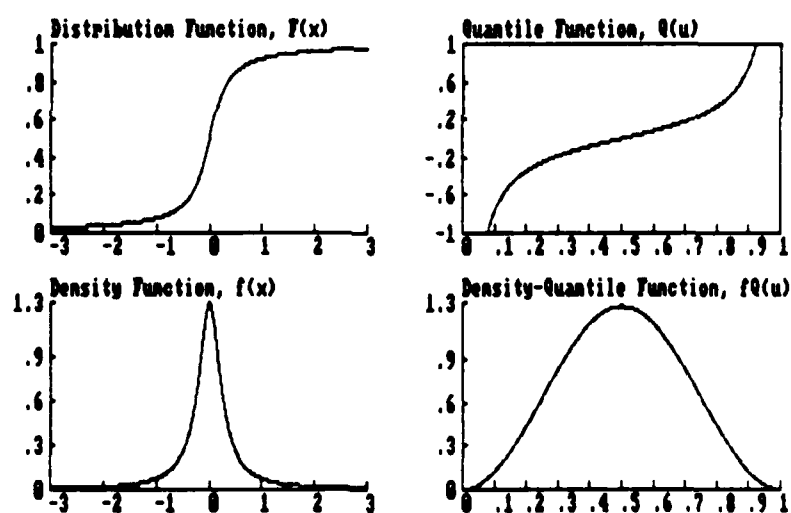


FIG. 19. Cauchy Distribution.

- The Pareto(p) Distribution, $p > 0$, is given by

$$F(x) = \begin{cases} 0, & x \leq (1 - 2^{1/p})/\sigma \\ 1 - [2^{1/p} + \sigma x]^{-p}, & x > (1 - 2^{1/p})/\sigma, \end{cases}$$

with quantile function

$$Q(u) = \frac{(1 - u)^{-1/p} - 2^{1/p}}{\sigma},$$

density function

$$f(x) = \begin{cases} 0, & x \leq (1 - 2^{1/p})/\sigma \\ \sigma p [2^{1/p} + \sigma x]^{-p-1}, & x > (1 - 2^{1/p})/\sigma, \end{cases}$$

and density-quantile function

$$fQ(u) = \sigma p (1 - u)^{(1/p)+1},$$

where $\sigma = 2 \cdot 4^{1/p} (1 - 3^{-1/p})$. These functions are plotted in Figure 20.

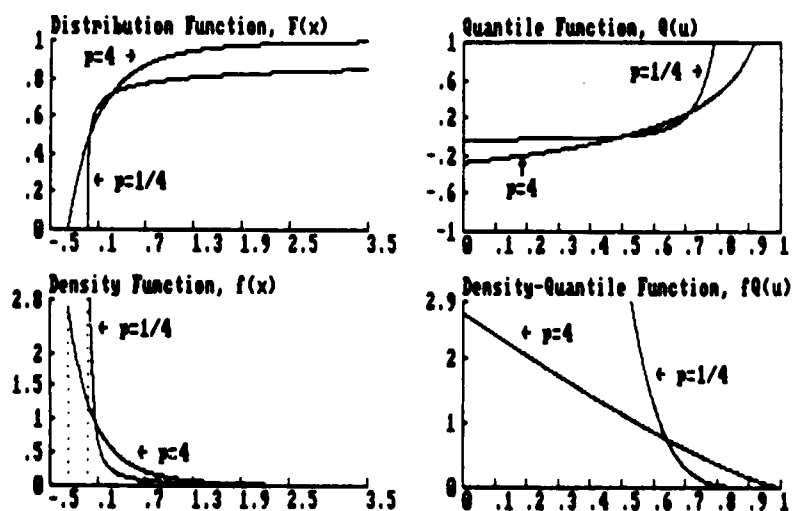


FIG. 20. Pareto(p) Distribution.

- The Fréchet(p) Distribution, $p > 0$, (which is also referred to as the Type II Extreme Value Distribution) is given by

$$F(x) = \begin{cases} 0, & x \leq -(\ln 2)^{-1/p}/\sigma \\ \exp \left\{ -[(\ln 2)^{-1/p} + \sigma x]^{-p} \right\}, & x > -(\ln 2)^{-1/p}/\sigma, \end{cases}$$

with quantile function

$$Q(u) = \frac{(-\ln u)^{-1/p} - (\ln 2)^{-1/p}}{\sigma},$$

density function

$$f(x) = \begin{cases} 0, & x \leq -(\ln 2)^{-1/p}/\sigma \\ \sigma p [(\ln 2)^{-1/p} + \sigma x]^{-p-1} \cdot \exp \left\{ -[(\ln 2)^{-1/p} + \sigma x]^{-p} \right\}, & x > -(\ln 2)^{-1/p}/\sigma, \end{cases}$$

and density-quantile function

$$fQ(u) = \sigma p u (-\ln u)^{(1/p)+1},$$

where $\sigma = 2[(\ln 4 - \ln 3)^{-1/p} - (\ln 4)^{-1/p}]$. These functions are plotted in Figure 21.

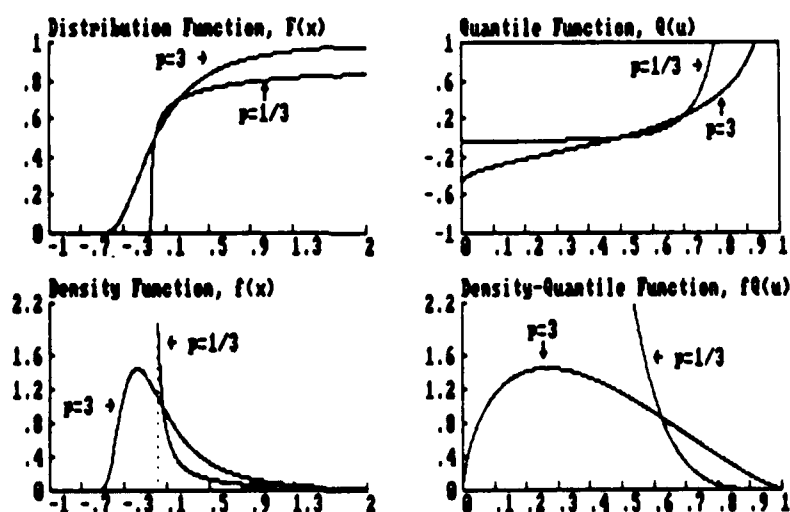


FIG. 21. Fréchet(p) Distribution.

APPENDIX B

PROOFS OF SECTION 2.3 THEOREMS

PROOF OF THEOREM 2.3.1. Consider

$$\begin{aligned}
 hQ(1-t)[Q(1-tu) - Q(1-t)] &= \frac{fQ(1-t)}{t} [Q(1-tu) - Q(1-t)] \\
 &= \frac{fQ(1-t)}{t} \int_{1-t}^{1-tu} q(z) dz \\
 &= \int_{1-t}^{1-tu} \frac{fQ(1-t)}{fQ(z)} \frac{1}{t} dz \\
 &= \int_u^1 \frac{fQ(1-t)}{fQ(1-tz)} dz \\
 &= \int_u^1 \frac{t^{\rho+1} L(t)}{(tz)^{\rho+1} L(tz)} dz \\
 &= \int_u^1 z^{-\rho-1} \frac{L(t)}{L(tz)} dz \\
 &= \int_u^1 z^{-\rho-1} dz \\
 &\quad + \int_u^1 z^{-\rho-1} \left[\frac{L(t)}{L(tz)} - 1 \right] dz \\
 &= -g(u; -\rho) + \epsilon(t, u, \rho).
 \end{aligned}$$

Since the quantile function of the exceedences over a threshold is

$$Q_{X-T | X>T}(u; T) = Q(1 - t^*(1 - u)) - Q(1 - t^*)$$

where $t^* = 1 - F(T)$, the theorem follows. \square

PROOF OF THEOREM 2.3.2(a). If $L(u)$ is a slowly varying function as $u \rightarrow 0^+$, then by Potter's Theorem, for any constants $A > 1$, $\alpha > 0$, there exists $T = T(A, \alpha)$ such that

$$\frac{L(t)}{L(tz)} \leq A \cdot \max\{z^{-\alpha}, z^{\alpha}\} \quad \text{for } 0 < t \leq T, 0 < z < 1.$$

Hence given $0 < \delta < 1$,

$$\left| z^{-\rho-1} \frac{L(t)}{L(tz)} \right| \leq Az^{-\rho-1} \cdot \max\{z^{-\alpha}, z^{\alpha}\},$$

which is an integrable function on $(\delta, 1)$.

Consider, for $\rho \in \mathbb{R}$,

$$\begin{aligned} \epsilon(t, u, \rho) &\leq \sup_{\delta < u < 1} \epsilon(t, u, \rho) \\ &= \sup_{\delta < u < 1} \int_u^1 z^{-\rho-1} \left[\frac{L(t)}{L(tz)} - 1 \right] dz \\ &= \int_{\delta}^1 z^{-\rho-1} \frac{L(t)}{L(tz)} dz - \int_{\delta}^1 z^{-\rho-1} dz \\ &\rightarrow \int_{\delta}^1 z^{-\rho-1} dz - \int_{\delta}^1 z^{-\rho-1} dz \quad \text{as } t \rightarrow 0^+ \\ &= 0. \end{aligned}$$

Thus, $\epsilon(t, u, \rho) \rightarrow 0$ uniformly in $\delta < u < 1$. \square

PROOF OF THEOREM 2.3.2(b). Under the assumption

$$\left| \frac{L(t)}{L(tu)} - 1 \right| \leq A(u)R(t)$$

for some positive measurable functions $A(u)$ and $R(t)$ where $\lim_{t \rightarrow 0^+} R(t) = 0$, then

$$\begin{aligned} |\epsilon(t, u, \rho)| &\leq \int_u^1 z^{-\rho-1} \left| \frac{L(t)}{L(tz)} - 1 \right| dz \\ &\leq \int_u^1 z^{-\rho-1} \cdot A(z)R(t) dz \\ &= R(t) \cdot \int_u^1 z^{-\rho-1} A(z) dz \\ &= A^*(u)R(t). \quad \square \end{aligned}$$

APPENDIX C

PROOF OF THEOREM 2.4.1

PROOF OF THEOREM 2.4.1. Since $fQ(1-u) = u^{\rho+1}L(u)$, where $\rho \in \mathbb{R}$, $\rho \neq 0$ and $L(u)$ is slowly varying as $u \rightarrow 0^+$, the conditional distribution of the exceedences can be written

$$Q_{X-T | X>T}(u; T) = \frac{1}{hQ(1-t^*)} [-g(1-u; -\rho) + \epsilon(t^*, 1-u, \rho)].$$

where $t^* = 1 - F(T)$.

Let $a = a(t^*)$, the scalar value of the function $a(\cdot)$. Then

$$\begin{aligned} E_{X-T | X>T} \left[1 + \frac{\rho}{a}(X-T) \right]^\alpha & \left[\ln \left\{ 1 + \frac{\rho}{a}(X-T) \right\} \right]^\beta \\ &= \int_0^1 \left[1 + \frac{\rho}{a} Q_{X-T | X>T}(u) \right]^\alpha \left[\ln \left\{ 1 + \frac{\rho}{a} Q_{X-T | X>T}(u) \right\} \right]^\beta du \\ &= \int_0^1 (1-u)^{-\rho\alpha} \frac{1}{[a(t^*)hQ(1-t^*)]^\alpha} \\ & \quad \cdot \left[1 + (1-u)^\rho \{ \rho\epsilon(t^*, 1-u, \rho) + [a(t^*)hQ(1-t^*) - 1] \} \right]^\alpha \\ & \quad \cdot \left\{ \ln(1-u)^{-\rho} - \ln a(t^*)hQ(1-t^*) \right. \\ & \quad \left. + \ln \left[1 + (1-u)^\rho \{ \rho\epsilon(t^*, 1-u, \rho) + [a(t^*)hQ(1-t^*) - 1] \} \right] \right\}^\beta du \\ &= \int_0^1 (1-u)^{-\rho\alpha} [\ln(1-u)^{-\rho}]^\beta \\ & \quad \cdot \frac{1}{[a(t^*)hQ(1-t^*)]^\alpha} \\ & \quad \cdot \left[1 + (1-u)^\rho \{ \rho\epsilon(t^*, 1-u, \rho) + [a(t^*)hQ(1-t^*) - 1] \} \right]^\alpha \\ & \quad \cdot \left(1 + \frac{1}{\ln(1-u)^{-\rho}} \right. \\ & \quad \left. \cdot \ln \left[\frac{1 + (1-u)^\rho \{ \rho\epsilon(t^*, 1-u, \rho) + [a(t^*)hQ(1-t^*) - 1] \}}{a(t^*)hQ(1-t^*)} \right] \right)^\beta du. \end{aligned}$$

If

$$\left| \frac{L(t)}{L(tu)} - 1 \right| \leq A(u)R(t),$$

then $|\epsilon(t, u, \rho)| \leq A^*(u)R(t)$. Also, it is assumed that $a(t^*)hQ(1 - t^*) = 1 + O(R_1(t^*))$. Hence,

$$\begin{aligned}
 & \left| \int_0^1 (1-u)^{-\rho\alpha} [\ln(1-u)^{-\rho}]^\beta du \right. \\
 & \quad \left. - E_{X-T | X>T} \left[1 + \frac{\rho}{a}(X-T) \right]^\alpha \left[\ln \left\{ 1 + \frac{\rho}{a}(X-T) \right\} \right]^\beta \right| \\
 &= \left| \int_0^1 (1-u)^{-\rho\alpha} [\ln(1-u)^{-\rho}]^\beta du \right. \\
 & \quad - \int_0^1 (1-u)^{-\rho\alpha} [\ln(1-u)^{-\rho}]^\beta \\
 & \quad \cdot \frac{1}{[a(t^*)hQ(1-t^*)]^\alpha} \\
 & \quad \cdot \left[1 + (1-u)^\rho \{ \rho\epsilon(t^*, 1-u, \rho) + [a(t^*)hQ(1-t^*) - 1] \} \right]^\alpha \\
 & \quad \cdot \left(1 + \frac{1}{\ln(1-u)^{-\rho}} \right. \\
 & \quad \cdot \ln \left[\frac{1 + (1-u)^\rho \{ \rho\epsilon(t^*, 1-u, \rho) + [a(t^*)hQ(1-t^*) - 1] \}}{a(t^*)hQ(1-t^*)} \right] \left. \right)^\beta du \left| \right. \\
 &\leq \left| \int_0^1 (1-u)^{-\rho\alpha} [\ln(1-u)^{-\rho}]^\beta du \right. \\
 & \quad - \int_0^1 (1-u)^{-\rho\alpha} [\ln(1-u)^{-\rho}]^\beta \\
 & \quad \cdot (1 + MR_1(t^*))^{-\alpha} \\
 & \quad \cdot (1 + (1-u)^\rho [\rho A^*(1-u)R(t^*) + MR_1(t^*)])^\alpha \\
 & \quad \cdot \left(1 - \frac{1}{\ln(1-u)^{-\rho}} \ln(1 + MR_1(t^*)) \right)^\beta \\
 & \quad \cdot \left(1 + \frac{1}{\ln(1-u)^{-\rho}} \right. \\
 & \quad \cdot \ln(1 + (1-u)^\rho [\rho A^*(1-u)R(t^*) + MR_1(t^*)]) \left. \right)^\beta du \left| \right. \\
 &\leq \left| \int_0^1 (1-u)^{-\rho\alpha} [\ln(1-u)^{-\rho}]^\beta du \right. \\
 & \quad - \int_0^1 (1-u)^{-\rho\alpha} [\ln(1-u)^{-\rho}]^\beta \\
 & \quad \cdot (1 + M_1 R_1(t^*)) \\
 & \quad \cdot (1 + M_2(1-u)^\rho [\rho A^*(1-u)R(t^*) + MR_1(t^*)])
 \end{aligned}$$

$$\begin{aligned}
& \cdot \left(1 + M_3 \frac{1}{\ln(1-u)^{-\rho}} \ln(1 + MR_1(t^*)) \right) \\
& \cdot \left(1 + M_4 \frac{1}{\ln(1-u)^{-\rho}} \right. \\
& \quad \left. \cdot \ln(1 + (1-u)^\rho [\rho A^*(1-u)R(t^*) + MR_1(t^*)]) \right) du \Big| \\
& \leq \left| M_1 R_1(t^*) \int_0^1 (1-u)^{-\rho\alpha} [\ln(1-u)^{-\rho}]^\beta du \right| \\
& \quad + \left| M_2 \int_0^1 (1-u)^{-\rho(\alpha-1)} [\ln(1-u)^{-\rho}]^\beta [\rho A^*(1-u)R(t^*) + MR_1(t^*)] du \right| \\
& \quad + \left| M_5 R_1(t^*) \int_0^1 (1-u)^{-\rho\alpha} [\ln(1-u)^{-\rho}]^{\beta-1} du \right| \\
& \quad + \left| M_6 \int_0^1 (1-u)^{-\rho(\alpha-1)} [\ln(1-u)^{-\rho}]^{\beta-1} \right. \\
& \quad \quad \left. [\rho A^*(1-u)R(t^*) + MR_1(t^*)] du \right| \\
& \leq M_7 R^*(t^*),
\end{aligned}$$

where $R^*(t^*) = \max\{R(t^*), R_1(t^*)\}$ and M_i are positive constants.

The other three expectations are found by changing the function $a(\cdot)$ and following the same arguments. \square

APPENDIX D

GRADIENT AND HESSIAN OF THE GPD LOG-LIKELIHOOD

Consider the space defined by $\mathcal{A} = \{-1 < \rho < 0, a > -\rho \cdot Y([nt_n]; [nt_n])\} \cup \{\rho > 0, a > 0\}$. On the space \mathcal{A} , the gradient vector of the GPD log-likelihood has elements

$$\begin{aligned} \frac{\partial \mathcal{L}_{\text{GPD}}(\rho, a; Y)}{\partial \rho} &= -\frac{[nt_n]}{\rho} \left(\frac{1}{\rho} + 1 \right) + \frac{1}{\rho^2} \sum_{i=1}^{[nt_n]} \ln \left(1 + \frac{\rho Y_i}{a} \right) \\ &\quad + \frac{1}{\rho} \left(\frac{1}{\rho} + 1 \right) \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a} \right)^{-1}, \\ \frac{\partial \mathcal{L}_{\text{GPD}}(\rho, a; Y)}{\partial a} &= \frac{[nt_n]}{\rho a} - \frac{1}{a} \left(\frac{1}{\rho} + 1 \right) \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a} \right)^{-1}. \end{aligned}$$

On the space \mathcal{A} , the Hessian matrix of the GPD log-likelihood has elements

$$\begin{aligned} \frac{\partial^2 \mathcal{L}_{\text{GPD}}(\rho, a; Y)}{\partial \rho^2} &= \frac{[nt_n]}{\rho^2} \left(\frac{3}{\rho} + 1 \right) - \frac{2}{\rho^3} \sum_{i=1}^{[nt_n]} \ln \left(1 + \frac{\rho Y_i}{a} \right) \\ &\quad - \frac{2}{\rho^2} \left(\frac{2}{\rho} + 1 \right) \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a} \right)^{-1} \\ &\quad + \frac{1}{\rho^2} \left(\frac{1}{\rho} + 1 \right) \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a} \right)^{-2}, \\ \frac{\partial^2 \mathcal{L}_{\text{GPD}}(\rho, a; Y)}{\partial a^2} &= -\frac{[nt_n]}{\rho a^2} + \frac{1}{a^2} \left(\frac{1}{\rho} + 1 \right) \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a} \right)^{-2}, \\ \frac{\partial^2 \mathcal{L}_{\text{GPD}}(\rho, a; Y)}{\partial \rho \partial a} &= -\frac{[nt_n]}{\rho^2 a} + \frac{1}{\rho a} \left(\frac{2}{\rho} + 1 \right) \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a} \right)^{-1} \\ &\quad - \frac{1}{\rho a} \left(\frac{1}{\rho} + 1 \right) \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a} \right)^{-2}. \end{aligned}$$

APPENDIX E

GRADIENT AND HESSIAN OF THE GEV LOG-LIKELIHOOD

Consider the space defined by $\mathcal{A} = \{-1 < \rho < 0, a > -\rho \cdot Y([nt_n]; [nt_n])\} \cup \{\rho > 0, a > 0\}$. On the space \mathcal{A} , the gradient vector of the GEV log-likelihood has elements

$$\begin{aligned} \frac{\partial \mathcal{L}_{\text{GEV}}(\rho, a; Y)}{\partial \rho} &= -\frac{[nt_n]}{\rho} \left(\frac{1}{\rho} + 1 \right) + \frac{1}{\rho^2} \sum_{i=1}^{[nt_n]} \ln \left(1 + \frac{\rho Y_i}{a} \right) \\ &\quad + \frac{1}{\rho} \left(\frac{1}{\rho} + 1 \right) \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a} \right)^{-1} \\ &\quad + \frac{1}{\rho^3} \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a} \right)^{-1/\rho} \ln \left(1 + \frac{\rho Y_i}{a} \right) \\ &\quad - \frac{1}{\rho^3} \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a} \right)^{-(1/\rho)+1} \ln \left(1 + \frac{\rho Y_i}{a} \right), \\ \frac{\partial \mathcal{L}_{\text{GEV}}(\rho, a; Y)}{\partial a} &= \frac{[nt_n]}{\rho a} - \frac{1}{a} \left(\frac{1}{\rho} + 1 \right) \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a} \right)^{-1} \\ &\quad + \frac{1}{\rho a} \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a} \right)^{-(1/\rho)-1} \\ &\quad - \frac{1}{\rho a} \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a} \right)^{-1/\rho}. \end{aligned}$$

On the space \mathcal{A} , the Hessian matrix of the GEV log-likelihood has elements

$$\begin{aligned} \frac{\partial^2 \mathcal{L}_{\text{GEV}}(\rho, a; Y)}{\partial \rho^2} &= \frac{[nt_n]}{\rho^2} \left(\frac{3}{\rho} + 1 \right) - \frac{2}{\rho^3} \sum_{i=1}^{[nt_n]} \ln \left(1 + \frac{\rho Y_i}{a} \right) \\ &\quad - \frac{2}{\rho^2} \left(\frac{2}{\rho} + 1 \right) \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a} \right)^{-1} \\ &\quad + \frac{1}{\rho^2} \left(\frac{1}{\rho} + 1 \right) \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a} \right)^{-2} \\ &\quad - \frac{1}{\rho^4} \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a} \right)^{-(1/\rho)-1} \end{aligned}$$

$$\begin{aligned}
& + \frac{2}{\rho^4} \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a}\right)^{-1/\rho} \\
& - \frac{1}{\rho^4} \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a}\right)^{-(1/\rho)+1} \\
& - \frac{3}{\rho^4} \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a}\right)^{-1/\rho} \ln \left(1 + \frac{\rho Y_i}{a}\right) \\
& + \frac{3}{\rho^4} \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a}\right)^{-(1/\rho)+1} \ln \left(1 + \frac{\rho Y_i}{a}\right) \\
& - \frac{1}{\rho^6} \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a}\right)^{-1/\rho} \left[\ln \left(1 + \frac{\rho Y_i}{a}\right) \right]^2 \\
& + \frac{2}{\rho^6} \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a}\right)^{-(1/\rho)+1} \left[\ln \left(1 + \frac{\rho Y_i}{a}\right) \right]^2 \\
& - \frac{1}{\rho^6} \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a}\right)^{-(1/\rho)+2} \left[\ln \left(1 + \frac{\rho Y_i}{a}\right) \right]^2,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}_{\text{GEV}}(\rho, a; Y)}{\partial a^2} = & - \frac{[nt_n]}{\rho a^2} + \frac{1}{a^2} \left(\frac{1}{\rho} + 1\right) \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a}\right)^{-2} \\
& - \frac{1}{\rho a^2} \left(\frac{1}{\rho} + 1\right) \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a}\right)^{-(1/\rho)-2} \\
& + \frac{2}{\rho^2 a^2} \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a}\right)^{-(1/\rho)-1} \\
& - \frac{1}{\rho a^2} \left(\frac{1}{\rho} - 1\right) \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a}\right)^{-1/\rho},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}_{\text{GEV}}(\rho, a; Y)}{\partial \rho \partial a} = & - \frac{[nt_n]}{\rho^2 a} + \frac{1}{\rho a} \left(\frac{2}{\rho} + 1\right) \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a}\right)^{-1} \\
& - \frac{1}{\rho a} \left(\frac{1}{\rho} + 1\right) \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a}\right)^{-2}
\end{aligned}$$

$$\begin{aligned}
& - \frac{1}{\rho^2 a} \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a}\right)^{-(1/\rho)-1} \\
& + \frac{1}{\rho^2 a} \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a}\right)^{-1/\rho} \\
& - \frac{1}{\rho^4 a} \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a}\right)^{-(1/\rho)-1} \ln \left(1 + \frac{\rho Y_i}{a}\right) \\
& + \frac{2}{\rho^4 a} \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a}\right)^{-1/\rho} \ln \left(1 + \frac{\rho Y_i}{a}\right) \\
& - \frac{1}{\rho^4 a} \sum_{i=1}^{[nt_n]} \left(1 + \frac{\rho Y_i}{a}\right)^{-(1/\rho)+1} \ln \left(1 + \frac{\rho Y_i}{a}\right).
\end{aligned}$$

APPENDIX F

DERIVATIVES OF HALL'S ESTIMATING EQUATIONS

For the estimating equations used for Hall's estimates for short tailed distributions, the derivatives are given by

$$\frac{\partial \psi_1(\rho, Q(1); X - T_n)}{\partial \rho} = -\frac{1}{\rho^2} \left(1 + \frac{1}{[nt_n]} \right) + \frac{2}{\rho^3} \ln \left(1 - \frac{X - T_n}{Q(1) - T_n} \right),$$

$$\begin{aligned} \frac{\partial \psi_2(\rho, Q(1); X - T_n)}{\partial Q(1)} &= \frac{1}{\rho} [Q(1) - T_n]^{-2} \\ &\quad \cdot \left\{ \left(1 + \frac{1}{[nt_n]} \right) - (\rho + 1) \left[1 - \frac{X - T_n}{Q(1) - T_n} \right]^{-2} \right\}, \end{aligned}$$

$$\begin{aligned} \frac{\partial \psi_1(\rho, Q(1); X - T_n)}{\partial Q(1)} &= \frac{\partial \psi_2(\rho, Q(1); X - T_n)}{\partial \rho} \\ &= \frac{1}{\rho^2} [Q(1) - T_n]^{-1} \left(1 - \left[1 - \frac{X - T_n}{Q(1) - T_n} \right]^{-1} \right). \end{aligned}$$

VITA

Scott D Grimshaw [REDACTED]
[REDACTED]
[REDACTED]

[REDACTED] He attended Southern Utah State College on academic scholarships and graduated Magna Cum Laude with a B.S. in Mathematics in June 1983. That fall he began graduate work in the Department of Statistics at Texas A&M University. He graduated in August 1985 with a M.S. in Statistics, writing his Master's thesis on *Estimation of the Linear-Plateau Segmented Regression Model in the Presence of Measurement Error*. From 1985-1986 he interned in the Statistics Internship Program, jointly sponsored by the Department of Statistics, Texas A&M University and the Applied Statistics Group, Engineering Department, E. I. du Pont de Nemours & Co. (Inc.). He was chosen as the 1987 W. S. Connor Memorial Award winner as the outstanding Ph.D. candidate in the Department of Statistics. At present, he is a candidate for the degree of Doctor of Philosophy. He has accepted a tenure track position beginning in the fall with the School of Business and Management, The University of Maryland, College Park. [REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]